



**HAL**  
open science

# Réseaux de régulation et éléments intégratifs et conjugatifs de la famille TnGBS dans l'adaptation de *Streptococcus agalactiae*

Violette da Cunha

► **To cite this version:**

Violette da Cunha. Réseaux de régulation et éléments intégratifs et conjugatifs de la famille TnGBS dans l'adaptation de *Streptococcus agalactiae*. Biologie cellulaire. Université Pierre et Marie Curie - Paris VI, 2012. Français. NNT : 2012PA066014 . tel-00828193

**HAL Id: tel-00828193**

**<https://theses.hal.science/tel-00828193>**

Submitted on 30 May 2013

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

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

**THESE DE DOCTORAT DE  
L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité

Ecole Doctorale Complexité du Vivant

Présentée par

M<sup>lle</sup>. DA CUNHA Violette

Pour obtenir le grade de

**DOCTEUR de l'UNIVERSITE PIERRE ET MARIE CURIE**

Sujet de la thèse :

Réseaux de régulation et éléments intégratifs et conjugatifs de la famille  
TnGBS dans l'adaptation de *Streptococcus agalactiae*

soutenue le 8 Mars 2012

devant le jury composé de :

M. DUJON Bernard	Directeur du jury
Mme. PETIT Marie-Agnès	Rapporteur
M. DEUTSCHER Josef	Rapporteur
Mme. COCAIGN-BOUSQUET Muriel	Examineur
M. LECLERCQ Roland	Examineur
M. GLASER Philippe	Directeur de thèse







## Sommaire

<b>1. INTRODUCTION.....</b>	<b>11</b>
<b>I. S. AGALACTIAE AU SEIN DU GENRE STREPTOCOQUE.....</b>	<b>12</b>
A. S. AGALACTIAE EST UNE BACTERIE PATHOGENE OPPORTUNISTE POUR L'HOMME ET L'ANIMAL. ....	14
1. Une bactérie pathogène de l'homme.....	14
2. Une bactérie commensale .....	16
3. La lutte contre S. agalactiae .....	17
4. Une bactérie pathogène chez l'animal.....	19
<b>II. LA DIVERSITE DES POPULATIONS.....</b>	<b>21</b>
A. DIFFERENTES APPROCHES POUR LE TYPAGE DE S. AGALACTIAE .....	21
B. LES APPORTS DU TYPAGE SUR L'ETUDE DE LA DIVERSITE .....	23
<b>III. LES FACTEURS DE VIRULENCE.....</b>	<b>25</b>
A. LES DIFFERENTES ETAPES DE L'INFECTION .....	25
1. Adhésion et colonisation des cellules épithéliales .....	25
2. Invasion et traversée des barrières de l'hôte.....	26
3. Evasion et échappement aux mécanismes de défenses de l'hôte.....	27
B. LES FACTEURS DE VIRULENCE CARACTERISES CHEZ S. AGALACTIAE .....	28
1. Les facteurs de virulence non protéiques.....	29
2. Les structures protéiques.....	35
3. L'opéron cyl.....	41
<b>IV. S. AGALACTIAE UNE BACTERIE RICHE EN SYSTEMES DE REGULATION.....</b>	<b>42</b>
A. LES SYSTEMES A DEUX COMPOSANTS.....	44
1. CovRS .....	44
2. DltRS .....	46
3. CiaRH .....	47
4. RgfBDAC ("Regulatory of Fibrinogen binding") .....	49
B. LES REGULATEURS SIMPLE (STAND ALONE) .....	50
1. RovS et les régulateurs de la famille Rgg.....	50
2. RogB.....	51
C. ADAPTATION A DIFFERENTS ENVIRONNEMENTS: ETUDES TRANSCRIPTOMIQUES.....	52
<b>V. ELEMENTS GENETIQUES MOBILES (EGMS) ET EVOLUTION DU GENOME CHEZ S. AGALACTIAE.....</b>	<b>54</b>
A. STRUCTURE COMPOSITE DU GENOME DE S. AGALACTIAE .....	54
B. LES ICES DES EGMS PARTICULIERS.....	55
C. MOBILITE DES ICES PAR CONJUGAISON .....	56
1. Description du processus de conjugaison .....	56
2. Le spectre de transfert.....	63
3. Immunité de conjugaison .....	63
4. Régulation du processus de conjugaison.....	64
<b>2. RESULTATS.....</b>	<b>69</b>

<b>I.</b>	<b>RESEAUX DE REGULATION ASSOCIES AUX COMPOSANTS DE LA PAROI .....</b>	<b>69</b>
A.	MATERIELS ET METHODES.....	70
1.	<i>Construction des mutants insertionnels des gènes <i>ItaS</i> et <i>gbs1493</i> .....</i>	70
2.	<i>Analyses transcriptomiques.....</i>	70
3.	<i>Classification des gènes (Clustering).....</i>	71
4.	<i>Recherche de motifs .....</i>	72
5.	<i>Effet de la colistine.....</i>	72
6.	<i>Analyse des données d'identification des sites d'initiation de la transcription (TSS).....</i>	73
B.	RESULTATS ET DISCUSSIONS.....	73
1.	<i>Effet sur le profil d'expression de perturbations touchant des structures non protéiques de la paroi (publication n°1) .....</i>	73
2.	<i>Réseaux de régulation impliquant des régulateurs de type RofA (publication n°2) .....</i>	88
3.	<i>Analyse combinée de l'ensemble des données de transcriptome.....</i>	94
<b>II.</b>	<b>ETUDE D'UNE NOUVELLE FAMILLE D'ICE, LES TNGBS .....</b>	<b>116</b>
A.	MATERIELS ET METHODES .....	116
1.	<i>Culture des souches .....</i>	116
2.	<i>Construction des mutants.....</i>	117
3.	<i>Expériences de conjugaison .....</i>	118
4.	<i>Extraction et analyse bioinformatique des séquences.....</i>	119
5.	<i>Effet de la mitomycine sur la production de la forme circulaire .....</i>	119
6.	<i>Activité du promoteur <math>P_{IR}</math>.....</i>	120
B.	RESULTATS ET DISCUSSION .....	120
1.	<i>Les TnGBS, une nouvelle famille d'ICE (publication n°3).....</i>	120
2.	<i>Analyse comparative de la diversité des ICEs de la famille TnGBS (publication n°4 et 5).....</i>	123
3.	<i>Données complémentaires de l'étude de la diversité des ICEs de la famille TnGBS.....</i>	129
4.	<i>Analyse du processus de conjugaison de TnGBS2 et sa régulation.....</i>	137
5.	<i>Gbs1141, Gbs1147 et le promoteur <math>P_{IR}</math>, trois éléments pouvant intervenir dans la régulation du processus de conjugaison.....</i>	141
	<b>DISCUSSION ET PERSPECTIVES.....</b>	<b>146</b>
<b>3.</b>	<b>BIBLIOGRAPHIE.....</b>	<b>152</b>
<b>4.</b>	<b>ANNEXES.....</b>	<b>183</b>

## Liste des Figures

Figure 1: Arbre phylogénétique du gène <i>recN</i> réalisé sur 65 espèces du genre streptocoque.....	13
Figure 2 : Evolution de l'incidence des infections néonatales à streptocoques du groupe B en France de 1997 à 2006 (réseau Epibac).....	18
Figure 3 : Visualisation par eBURST des complexes clonaux reflétant la diversité au sein de la population de <i>S. agalactiae</i> .....	23
Figure 4 : Paroi de <i>S. agalactiae</i> .....	29
Figure 5: Comparaison et organisation du locus capsulaire .....	30
Figure 6: Structure de l'antigène B.....	31
Figure 7 : Structure et D-alanylation du LTA.....	34
Figure 8: Organisation du locus <i>secA2</i> dans les souches NEM316 et COH1 de <i>S. agalactiae</i> ..	39
Figure 9 : Loci pili chez NEM316.....	40
Figure 10: Organisation de l'opéron <i>cyl</i> chez <i>S. agalactiae</i> .....	42
Figure 11: Les systèmes à deux composants et voies de transduction du signal.....	43
Figure 12 : Organisation de l'opéron <i>dlt</i> au sein du genre streptocoque .....	46
Figure 13 : Structure secondaire et localisation de <i>csRNA</i> chez <i>S. agalactiae</i> .....	49
Figure 14: Différentes étapes du processus de conjugaison des ICEs. ....	57
Figure 15 : Composants de l'appareil de transfert des plasmides <i>pCF10</i> et <i>pIP501</i> et des deux ICEs <i>ICEBs1</i> et <i>Tn916</i> . ....	60
Figure 16: Effet des différentes concentrations de colistine sur la croissance de la souche NEM316 .....	80
Figure 17 Effets de doses sub-inhibitrices de colistine sur la croissance de la souche NEM316.	80
Figure 18: Photographie en microscopie optique du mutant insertionnel de l'antigène B et de la souche sauvage. Le mutant insertionnel du gène <i>gbs1493</i> correspond à l'image A, et la souche NEM316 sauvage correspond à l'image B. ....	88
Figure 19: Modèle d'action de <i>Rga</i> au sein des réseaux de régulation de <i>S. agalactiae</i> .....	91
Figure 20: Région promotrice des gènes divergents <i>srr1</i> et <i>rga</i> .....	92
Figure 21: Classification hiérarchique de 636 gènes. ....	101
Figure 22: Le cluster <i>CovR</i> . ....	103
Figure 23: Les clusters <i>CiaRH</i> , <i>Rga</i> , <i>HrcA</i> , <i>GlnR</i> , <i>GntR</i> et <i>Rex</i> . ....	105
Figure 24: Les clusters 7, 24 et 11.....	111
Figure 25: Clusters 9 et 22. ....	112
Figure 26: Expression des gènes de <i>TnGBS2</i> , analysée par microarrays, dans NEM316.....	129
Figure 27: Alignement des protéines apparentées à <i>Gbs1141</i> .....	130
Figure 28: Localisations des sites d'initiation de la transcription identifiés pour <i>TnGBS1</i> et <i>TnGBS2</i> .....	132
Figure 29: Localisation des petits ARNs non codants potentiels de <i>TnGBS1</i> et de <i>TnGBS2</i> .....	133
Figure 30: Alignement de la région localisée entre l'IR droite et la transposase et contenant deux petits ARNs non codants.....	135
Figure 31: $P_{IR}$ un promoteur également conservé dans les IRs des ISS, et des ICEs apparentés à <i>TnGBS2</i> . ....	137
Figure 32: Immunité de conjugaison de <i>TnGBS2</i> .....	140



*Figure 33: Orientation des promoteurs formés par les IRs dans les constructions  $P_{IR}$  et  $P_{IRinv}$*   
..... 144

## Liste des Tables

Tableau 1: Etudes MLST publiées et origines géographiques des souches.....	22
Tableau 2: Liste des régulateurs transcriptionnels de la souche NEM316 de <i>S. agalactiae</i> .....	44
Tableau 3: Liste des analyses transcriptomiques réalisées dans <i>S. agalactiae</i> .....	53
Tableau 4: Les différentes étapes du processus de conjugaison de pAD1 et pCF10 ainsi que les protéines impliquées dans ces différentes étapes. ....	68
Tableau 5: Liste des oligonucléotides utilisés pour la première partie de la thèse.....	70
Tableau 6: Gènes différentiellement exprimés en conditions d'adaptation à l'ajout de colistine. 83	
Tableau 7: Ratio observé pour les gènes du locus pilus PI-2a, <i>srr1</i> , <i>rogB</i> , <i>rga</i> , et <i>gbs1605-06</i> dans les analyses transcriptomiques des mutants $\Delta$ <i>rga</i> , $\Delta$ <i>rogB</i> , $\Delta$ <i>CovRS</i> .....	90
Tableau 8: Liste des analyses transcriptomiques réalisées dans la souche NEM316.....	98
Tableau 9: Identification de sites de fixation potentiels de <i>GntR</i> .....	107
Tableau 10: Identification de sites de fixation potentiels de <i>GlnR</i> .....	108
Tableau 11: Identification de sites de fixation potentiels de <i>HrcA</i> .....	109
Tableau 12: Identification de sites de fixation potentiels de <i>Rex</i> .....	110
Tableau 13: Liste des primers utilisés pour la construction des mutants et des différents plasmides utilisés dans la 2ème partie de la thèse.....	118
Tableau 14: Analyse de la production de formes circulaires en présence ou absence de bactéries réceptrices. ....	138
Tableau 15: Quantification de la production de formes circulaires de TnGBS2 en réponse à différentes concentrations de mitomycine C. ....	139
Tableau 16: Quantification de la production de formes circulaires de TnGBS2 dans le mutant $\Delta$ <i>gbs1141</i> .....	142
Tableau 17: Quantification de la production de formes circulaires de TnGBS2 dans le mutant $\Delta$ <i>gbs1147</i> .....	143
Tableau 18: Activité bêta-galactosidase du promoteur $P_{IR}$ de TnGBS2. ....	144

## Liste des Abréviations

ADN	Acide DésoxyriboNucléique
ARN	Acide RiboNucléique
ARNr	Acide RiboNucléique ribosomal
CAMP	Initiales des nom des auteurs qui ont décrit le test en 1944 <u>C</u> hrisie, <u>A</u> tkins, et <u>M</u> unch- <u>P</u> eterson
CC	<u>C</u> omplexes <u>C</u> lonaux
CGH	Hybridation de Génome Complet ( " <u>C</u> omplet <u>G</u> enome <u>H</u> ybridization")
CNR	Centre National de Référence
FH	Facteur H
GBS	Streptocoque du Groupe B ( " <u>G</u> roup <u>B</u> <u>S</u> treptococcus")
ICE	Élément Intégratif et Conjugatif ( " <u>I</u> ntegrative and <u>C</u> onjugative <u>E</u> lement")
Ig	Immunoglobulines
INP	<u>I</u> nfection <u>N</u> éonatale <u>P</u> récoce
INT	<u>I</u> nfection <u>N</u> éonatale <u>T</u> ardive
IR	Répétitions Inverses ( <u>I</u> nverted <u>R</u> epet)
IS	Séquence d'Insertion ( <u>I</u> nsertion <u>S</u> equences)
LTA	Acide <u>L</u> ipotéichoïque
MEC	<u>M</u> atrice <u>E</u> xtracellulaire
MLST	Schéma de séquençage de gènes multiples ( <u>M</u> ulti <u>L</u> ocus <u>S</u> equences <u>T</u> yping)
PBP	( <u>P</u> enicilline <u>B</u> inding <u>P</u> rotein)
PCR	Réaction de polymérisation en Chaîne (Polymerase Chain Reaction)
SA	<u>S</u> emaine d' <u>A</u> ménorrhée
ST	<u>S</u> equences <u>T</u> ype
TCS	Système à deux composants ( " <u>T</u> wo <u>C</u> omponent <u>S</u> ystem")
TH	Milieu <u>T</u> odd <u>H</u> ewith
TSS	Site d'initiation de la transcription ( " <u>T</u> ranscription <u>S</u> tart <u>S</u> ite")
T4SS	Système de sécrétion de type IV ( " <u>T</u> ype <u>I</u> V <u>S</u> ecretion <u>S</u> ystem")

# 1. INTRODUCTION

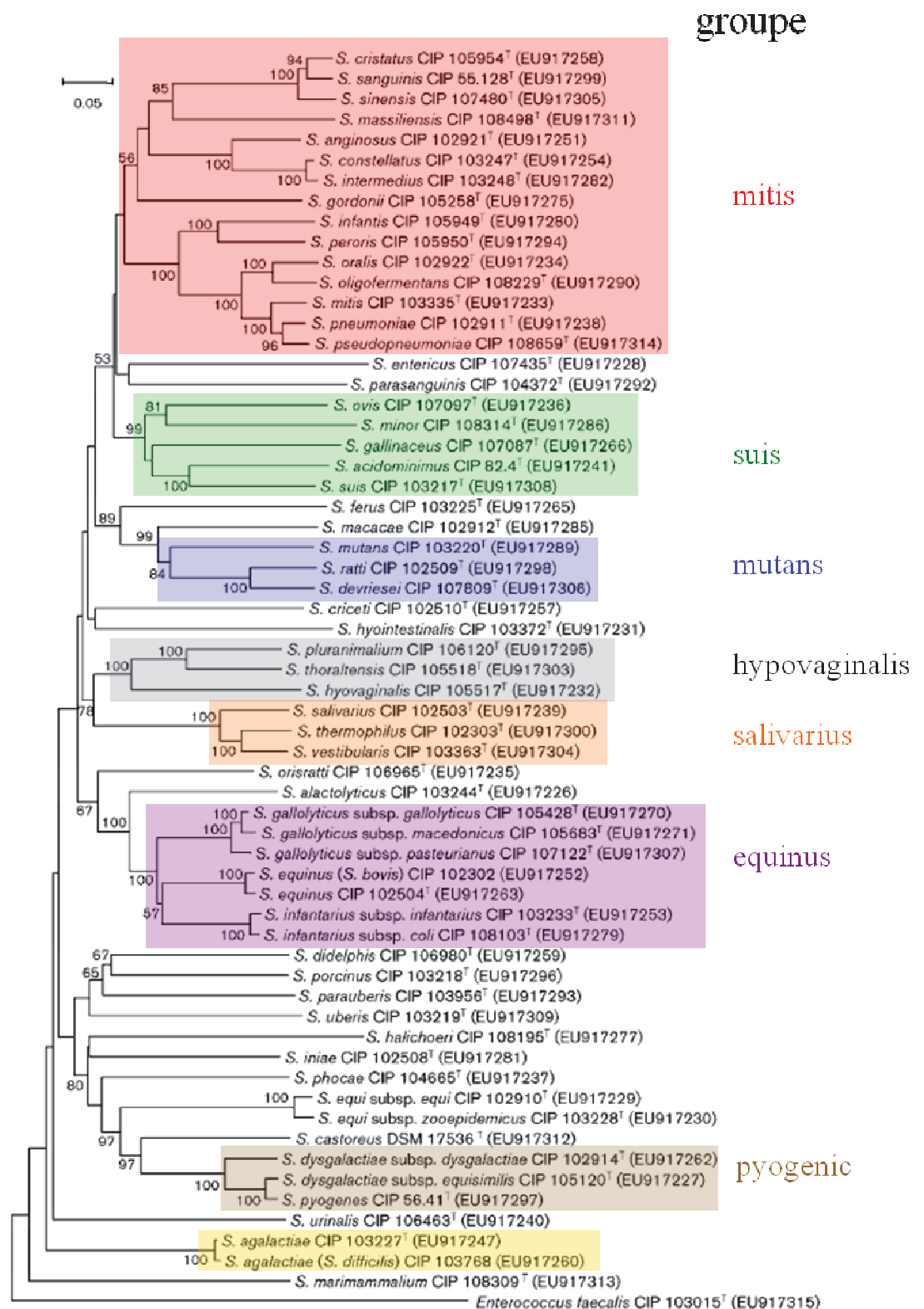
*Streptococcus agalactiae*, également appelé Streptocoque du Groupe B ou GBS, est la première cause de septicémies et de méningites chez les nouveaux-nés, et est également responsable d'infections chez les personnes âgées et les personnes immunodéprimées. On le retrouve aussi associé à des infections chez l'animal. Néanmoins, c'est avant tout une bactérie commensale du tube digestif présente chez 20 % de la population humaine. *S. agalactiae* est retrouvé dans différents hôtes, il a donc été capable de s'adapter au cours de son évolution à ces différents environnements. Pour s'adapter à de nouveaux environnements ou pour répondre à des modifications de son environnement usuel, une bactérie peut mettre en place différentes stratégies. Tout d'abord la bactérie peut modifier son contenu génétique. L'acquisition ou la modification de facteurs par transferts horizontaux est possible grâce à la mobilisation d'éléments génétiques mobiles (EGMs), qui sont connus pour être un des moteurs de l'évolution des génomes bactériens. Une autre stratégie est la mise en place par la bactérie de systèmes de régulation, qui conduisent à une réponse adaptée de la bactérie à des modifications de son environnement, en régulant l'expression de gènes, lui permettant de se maintenir et de se propager dans cet environnement modifié. Les systèmes à deux composants (TCS) sont largement répandus chez les bactéries et constituent la principale stratégie mise en place par celles-ci pour percevoir et répondre aux modifications de leur environnement.

Les deux parties de mon travail de thèse portent, d'une part sur l'analyse de réseaux de régulation associés à la perception de différents environnements rencontrés par la bactérie. La seconde partie de mon travail porte sur l'analyse chez *S. agalactiae* de la régulation du transfert, du maintien et de la diversité des éléments intégratifs et conjugatifs (ICEs) de la famille TnGBS. Cette famille d'ICEs est impliquée dans des phénomènes de transferts horizontaux de grands fragments d'ADN chromosomique.

## I. *S. agalactiae* au sein du genre streptocoque

*S. agalactiae* est une bactérie Gram positif à bas GC% appartenant à la famille des *Streptococcaceae*, qui regroupe les genres *Streptococcus*, *Enterococcus* et *Lactococcus*. Ces trois genres partagent les mêmes caractéristiques: cocci à Gram positif, non sporulés, à métabolisme aéro-anaérobie facultatif et dépourvus de catalase. Au moment de la rédaction de cette thèse, le genre streptocoque comporte 95 espèces et 17 sous-espèces. L'ensemble des espèces est référencé sur le site <http://www.bacterio.cict.fr/s/streptococcus.html>. De nouvelles espèces sont continuellement isolées d'échantillons humains, animaux et environnementaux. Les bactéries du genre streptocoque sont des bactéries commensales humaines et animales, de pathogénicité variable. Deux espèces sont utilisées en agroalimentaire, *S. thermophilus* et *S. macedonicus* appartenant respectivement au groupe salivarius et bovis. L'espèce *S. agalactiae* a initialement été décrite chez les bovins, par Nocard et Mollereau en 1887, comme le "streptocoque des mammites bovines"[1]. Le nom actuel d'*agalactiae*, qui fait allusion à une absence de lait, date de 1896 et a été donné par Lehrmann et Neumann [2].

La classification des streptocoques a d'abord été fondée sur l'étude de leurs caractères phénotypiques, puis sur l'étude de leurs caractères génotypiques. En 1933, Rebecca Lancefield proposa la première classification moderne des streptocoques [3]. Cette classification est basée sur les propriétés antigéniques d'un composé pariétal des streptocoques, le polyside C. Dans cette classification, *S. agalactiae* a été désigné comme un streptocoque du groupe B (GBS), et correspond en fait à l'unique espèce de ce groupe [4]. Une étude phylogénétique récente [5], portant sur l'analyse d'une séquence partielle du gène *recN* (1249 pb) chez 65 espèces et 12 sous-espèces, définit au sein du genre streptocoque sept groupes distincts: salivarius, equinus, hyovaginalis, pyogenes, mutans, suis et mitis (Figure 1). Le groupe mitis lui-même est constitué de trois sous-groupes: le sous-groupe sanguinis (*S. sanguinis*/*S. cristatus*/*S. sinensis*), le sous-groupe anginosus (*S. anginosus*/*S. intermedius*/*S. constellatus*) et le sous-groupe mitis (*S. mitis*/*S. infantis*/*S. peroris*/*S. oralis*/*S. oligofermentans*/*S. pneumoniae*/*S. pseudopneumoniae*). Cette analyse est largement concordante avec des études phylogénétiques réalisées à partir de la séquence de l'ARNr 16S [6]. Dans cette analyse *S. agalactiae* n'est situé dans aucun de ces groupes. Cependant, différentes données montrent que cette espèce est phylogénétiquement plus proche des espèces appartenant au groupe pyogenes.



**Figure 1:** Arbre phylogénétique du gène *recN* réalisé sur 65 espèces du genre streptocoque

Arbre réalisé en utilisant la méthode de neighbour-joining, la taille de l'échelle indique 0.05 changement par position nucléotidique. La séquence d'*Enterococcus faecalis* est utilisée comme groupe externe. Les valeurs de Bootstrap supérieures à 50 pour 100 réplicats sont indiquées. Les différents groupes sont indiqués par différentes couleurs. D'après [5].

## **A. *S. agalactiae* est une bactérie pathogène opportuniste pour l'homme et l'animal.**

### **1. Une bactérie pathogène de l'homme**

#### **a) Infections néonatales**

Les premiers cas d'infections néonatales ont été décrits par Eickhoff en 1964 [7]. Cette identification, relativement récente, laisse penser à une émergence de cas pathologiques dans les années 60-70 [8, 9]. Aujourd'hui, malgré les procédures de prévention mises en place, *S. agalactiae* reste la première cause bactérienne d'infections néonatales en France devant *E. coli*, en terme de prévalence [10]. Ces infections se traduisent généralement par des pneumonies, des septicémies ou des méningites pouvant entraîner des séquelles neurologiques graves et dans certains cas le décès. Les infections néonatales sont classées en syndromes précoces, et tardifs en fonction de l'âge du nouveau-né.

#### **(1) Syndrome précoce**

Les infections néonatales précoces (INP) se manifestent jusqu'au septième jour de vie et principalement au cours des premières heures, et se présentent souvent comme des pneumonies, ou des détresses respiratoires, qui peuvent évoluer rapidement en bactériémie et dans les cas les plus graves en choc septique. L'incidence des INP est de 3/10000 naissances et ces infections sont létales dans 2,5% des cas [11], soit 210 cas d'infections et 5-6 décès pour 700 000 naissances annuelles en France. Lors des infections précoces, les bactéries sont transmises de la mère à l'enfant par la contamination du liquide amniotique après rupture de la membrane placentaire, ou lors de l'accouchement par aspiration des fluides vaginaux infectés. 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 [12].

## **(2) Syndrome tardif**

Les infections néonatales tardives (INT) peuvent apparaître jusqu'à sept mois après la naissance et se distinguent des infections précoces par une bactériémie avec un fort taux de progression en méningite [13]. En France, une étude récente du Centre National de Référence (CNR) indique que ces infections sont généralement associées à des septicémies, dans 65% des cas à des méningites [11], et sont létales dans 14% des cas. Les enfants qui survivent à une méningite à streptocoques B, peuvent dans 25 à 35% des cas, souffrir à long terme de séquelles neurologiques permanentes (pertes auditives, et/ou troubles cognitifs). L'incidence des INT est de 1/10000 naissances et ces infections représentent en France en moyenne 70 cas par an [14].

Le mode de transmission et le processus de l'infection restent incertains, bien que dans certains cas une transmission de la mère à l'enfant ait été démontrée. Un scénario possible implique une colonisation intestinale précoce, qui conduirait dans les premiers jours de la vie à la multiplication de la bactérie dans la lumière intestinale. La translocation de la bactérie à travers l'épithélium intestinal conduirait à la dissémination de l'infection par la circulation sanguine, cette dissémination aboutissant en méningite en cas de traversée de la barrière hémato-encéphalique [15]. La colonisation intestinale des nouveaux-nés est accréditée par différentes observations :

(a) 40 à 60 % des nouveaux-nés colonisés de manière asymptomatique par GBS à la naissance sont également positifs au niveau du rectum à 4 et 12 semaines de vie [16]

(b) une étude portant sur la colonisation vaginale et rectale montre que les femmes sont colonisées par un seul clone, lequel est également fréquemment trouvé dans les *fèces* du nouveau-né pendant un an [17].

### **b) *S. agalactiae* est également un pathogène chez les adultes**

En 1940, une première étude rapporte deux cas d'arthrites chez des personnes âgées [18]. Depuis les années 2000 de plus en plus d'études rapportent des cas d'infections à *S. agalactiae* chez des adultes [19, 20]. Les manifestations cliniques sont extrêmement diverses notamment: des septicémies sans foyers identifiés, des infections de la peau, des infections des voies



respiratoires, des arthrites, des ostéomyélites, des méningites, ainsi que des infections des voies urogénitales [21, 22]. Ces infections sont plus fréquentes chez les individus âgés de plus de 65 ans, ainsi que chez les sujets affaiblis (dénutrition, diabète, cirrhose, insuffisance rénale, cancers...) [23, 24]. Une étude rapporte que chez les adultes de plus de 65 ans les infections invasives à GBS ont une incidence de 26/100 000 personnes aux Etats-Unis [25]. Le taux de mortalité est élevé et varie selon les études entre 18 et 25%. Les sérotypes capsulaires Ia, III et V représentent 70% des sérotypes isolés, mais sont également ceux que l'on retrouve le plus souvent chez les personnes âgées colonisées et en bonne santé [22, 26].

Chez la femme enceinte ou en post-partum, l'infection peut conduire à des endométrites et à une stérilité. Si l'impact de la colonisation vaginale maternelle sur les infections néonatales à *S. agalactiae* a été largement étudié, peu d'études concernent 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 [27, 28].

## 2. Une bactérie commensale

*S. agalactiae* est considéré comme commensal du tube digestif de l'homme, le plus souvent en faible concentration avec une possibilité de contamination de la sphère urogénitale. En France le portage génital est d'environ 23% au sein de la population adulte [29]. Cependant des souches de *S. agalactiae* ont également été retrouvées sur la peau ou dans la gorge [29]. La colonisation est généralement asymptomatique et seuls des examens bactériologiques permettent d'identifier les porteurs. Le portage de cette bactérie apparaît comme dynamique. Une étude danoise, réalisée sur des femmes enceintes pendant une période de deux ans, montre que pour environ 50% des femmes, une détection intermittente du même clone a été constatée [17]. Cette étude démontre que le portage n'est pas toujours repérable, et est donc sous-estimé du fait de la limite de sensibilité des méthodes d'identification par culture ou par PCR. Ainsi dans les cas de portage intermittent, les tests négatifs pourraient correspondre à des concentrations bactériennes inférieures au seuil de sensibilité.

Une étude du portage chez différents membres d'une famille, portant sur 37 couples, propose que le contact sexuel soit un facteur important de la transmission de *S. agalactiae*. Pour 10 des 11 couples dont les deux partenaires sont porteurs, les séquences type (ST) des souches isolées chez les deux partenaires étaient identiques [29]. Cette transmission par contact sexuel

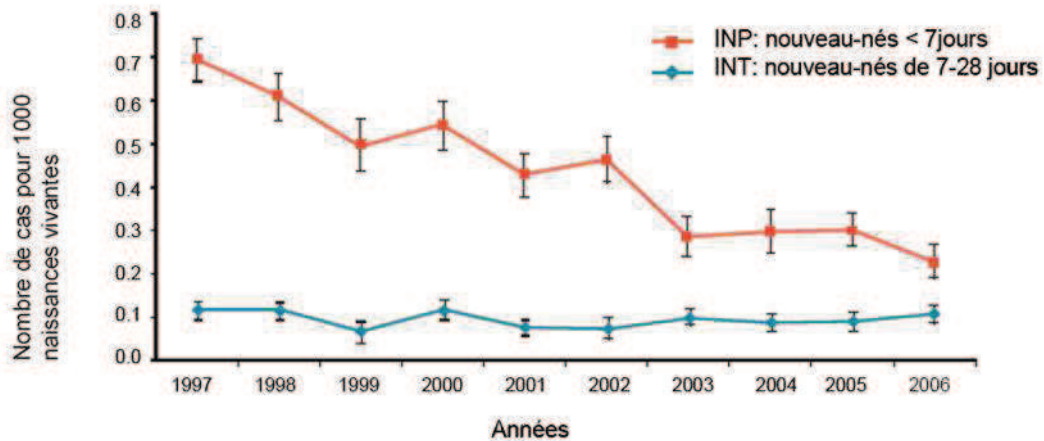
avait également été suggérée dans des études précédentes [16, 30]. Dans cette étude, aucune relation n'a été observée entre les souches portées par les parents et celles portées par les enfants, suggérant qu'il n'y aurait pas de transmission de *S. agalactiae* des parents aux enfants.

### 3. La lutte contre *S. agalactiae*

#### a) Prévention des infections néonatales

En 2001, dans le cadre de la prévention anténatale du risque d'infection du nouveau-né, un dépistage systématique du portage vaginal de *S. agalactiae* en fin de grossesse, entre la 33<sup>ème</sup> et la 39<sup>ème</sup> semaine d'aménorrhée a été mis en place suite aux recommandations de l'ANAES (Agence Nationale d'Accréditation et d'Evaluation en Santé). Ce dépistage a permis de montrer que la prévalence chez les femmes enceintes en France atteint 23%, soit au moins 161 000 femmes/an. Les valeurs prédictives d'un dépistage prénatal sur la colonisation au moment de l'accouchement varient inversement avec le temps écoulé entre le dépistage et l'accouchement. En cas de dépistage positif au cours de la grossesse, une antibioprophylaxie per-partum est mise en place; celle-ci consiste en des injections par voie intra-veineuse de pénicilline G ou d'amoxicilline toutes les 4h dès le déclenchement du travail jusqu'à l'expulsion. En cas d'allergie à la pénicilline, l'administration d'érythromycine IV ou d'une céphalosporine est recommandée en France. Aucune souche de *S. agalactiae* résistante à la pénicilline n'a encore été décrite [31]. Par contre les données du CNR sur *S. agalactiae* rapportent une augmentation du taux de résistance à l'érythromycine: ainsi 13,8% des souches isolées chez les nourrissons atteints d'infections néonatales invasives en 2008 ont été détectées comme résistantes à l'érythromycine.

La mise en place du dépistage et de l'antibioprophylaxie a permis de diminuer par un facteur trois, l'incidence des infections précoces (Figure 2). Par contre elle n'a pas eu d'impact sur les infections tardives. Une étude de 2008 permet d'expliquer la persistance à faible fréquence des infections précoces par une absence de dépistage maternel, et/ou une réalisation trop précoce du dépistage au cours de la grossesse [11]. Yancey et al. ont montré en effet que 4% des femmes avec un dépistage négatif entre la 33<sup>ème</sup> et la 39<sup>ème</sup> semaine d'aménorrhée (SA) étaient porteuses de GBS au moment de l'accouchement [32].



**Figure 2 :** Evolution de l'incidence des infections néonatales à streptocoques du groupe B en France de 1997 à 2006 (réseau Epibac)  
D'après [14]

## b) Développement d'un vaccin

Un complément au dépistage systématique associé à une antibioprophylaxie serait la vaccination des femmes avant ou pendant la grossesse. L'objectif de ce vaccin serait de transmettre des anticorps protecteurs au nouveau-né. Une vaccination efficace doit prévenir à la fois les syndromes précoces et tardifs. La combinaison d'antigènes doit pouvoir protéger de toutes les souches de *S. agalactiae*. En outre, un vaccin permettrait de réduire les risques liés aux allergies à la pénicilline et à l'émergence de souches résistantes aux antibiotiques.

Les bases du développement de ce vaccin reposent sur différentes observations : (a) les mères dont les enfants ont développé une infection précoce, présentent un taux d'IgG sérotype-spécifiques en général plus bas que celui de mères colonisées dont les enfants n'ont pas développé d'infections, et (b) les enfants infectés présentent un taux d'IgG plus bas que ceux non infectés [33, 34]. Ces études ont permis d'établir qu'un taux élevé d'IgG sérotype-spécifiques chez la mère permettrait une protection de l'enfant [35]. Les premiers essais de vaccins dans les années 1980 ont été réalisés par l'injection de polysides capsulaires purifiés [35], la capsule étant le facteur de virulence majeur de *S. agalactiae*. Afin d'améliorer la réponse spécifique, des vaccins conjugués contre neuf sérotypes ont été testés et se sont révélés immunogènes chez la souris, le lapin et le babouin [36]. Peu ou aucune protection croisée n'a été observée entre les différents sérotypes et le développement de vaccins multivalents s'est donc avéré nécessaire. Pour couvrir 95% de la population d'Europe et d'Amérique du Nord le vaccin devrait être dirigé

contre au moins cinq sérotypes (Ia, Ib, II, III, et V). Les stratégies vaccinales basées sur l'immunisation par les polysides capsulaires montrent leurs limites du fait de la diversité de ces derniers.

Récemment, la stratégie de vaccinologie inverse, d'abord utilisée pour lutter contre les souches de *Neisseria meningitidis* de sérotype B [37], a été également appliquée pour développer un vaccin contre *S. agalactiae* [38]. Cette stratégie repose sur l'identification des candidats vaccinaux par des analyses génomiques. Une protéine candidate doit être exposée à la surface de la bactérie et être immunogène. La capacité à conférer une immunité protectrice chez la souris a été testée par une immunisation de la mère ou par une immunisation du souriceau. Par cette approche quatre candidats ont été retenus: la protéine sécrétée Sip et trois protéines à motif LPXTG (Sag1408, Sag0645, et Sag0649) qui sont des composants de deux structures de type pilus [38]. Cependant, parmi ces quatre protéines, Sip est la seule à être conservée dans les huit souches séquencées de GBS, mais la protection par des anticorps contre cette protéine seule est limitée. Les souches peuvent échapper à ce vaccin soit en exprimant d'autres types de pilus, soit en réprimant l'expression du pilus, le pilus n'étant pas essentiel à la virulence. Cette approche n'a pas permis d'identifier d'antigènes universels, ce qui peut être lié à la diversité des protéines de surface de *S. agalactiae* [39].

La détermination d'antigènes pour un vaccin universel contre *S. agalactiae* est difficile pour au moins trois raisons: i) *S. agalactiae* exprime un vaste répertoire de structures de surface, la présentation à la surface de ces structures peut être variable soit du fait de leur niveau d'expression soit de leur accessibilité [40]. ii) Certains polymorphismes peuvent être à l'origine d'une faible protection croisée. iii) Le vaccin doit agir au niveau du nouveau-né et non sur le portage maternel, afin d'éviter des remplacements de souches au niveau des populations de portage.

#### **4. Une bactérie pathogène chez l'animal**

*S. agalactiae* est également une bactérie pathogène pour les animaux, en particulier pour les bovins et les poissons.

## a) Chez les bovins

*S. agalactiae* est une cause importante de mammites dans les élevages laitiers, conduisant à une diminution de la quantité et de la qualité du lait. Souvent l'infection reste limitée à la partie plus externe du système des canaux lactifères. Ce n'est qu'après de longues périodes d'infections que les bactéries se répandent plus profondément dans le pis. L'infection devient ensuite chronique et conduit à un arrêt de la lactation. Les programmes de thérapie de tous les individus infectés d'un troupeau sont en général efficaces pour éradiquer le pathogène d'un troupeau, spécialement s'ils sont accompagnés de bonnes techniques d'hygiène de la traite [41]. Une forte association entre la capacité des souches à fermenter le lactose et leur origine a été observée; une étude montre que 92% des souches d'origine bovine fermentent le lactose contre seulement 13% des souches d'origine humaine [42]. Le lactose correspondant approximativement à 5% du poids du lait [43], cette aptitude est un moyen d'adaptation de *S. agalactiae* à l'environnement du lait et des canaux lactifères. Mise à part cette activité métabolique, l'adaptation de *S. agalactiae* à l'environnement des canaux lactifères n'a pas été très étudiée. Il y a probablement une transmission de vache à vache lors de la traite. Le portage bovin au niveau du tube digestif n'a pas été étudié systématiquement.

## b) Chez le poisson

Avant le milieu des années 1980, quelques articles ont fait état de l'isolement de souches de *S. agalactiae* chez des poissons, et une étude a signalé que la bactérie semble être à l'origine d'une mortalité importante dans des élevages de l'Alabama [44]. A partir de 1984, les poissons d'élevages israéliens de *Tilapia* sp., et de truite arc-en-ciel (*Oncorhynchus mykiss*) ont été victimes de méningo-encéphalites dues à des streptocoques et provoquant la mort de 30 à 50% des animaux. Les analyses microbiologiques ont permis d'isoler, à partir du cerveau des tilapias ou des truites, des reins des tilapias ou du sang des truites, des souches de streptocoques dont les caractères biochimiques, sérologiques et génétiques (hybridation ADN - ADN, détermination du G + C %) suggéraient qu'elles constituaient deux nouvelles espèces, *Streptococcus difficile* et *Streptococcus shiloi* [45]. Ultérieurement, les souches de *Streptococcus shiloi* ont été assimilées à *Streptococcus iniae* et celles de *Streptococcus difficile* à *S. agalactiae* [46]. *Streptococcus*

*agalactiae* est capable d'infecter, outre les tilapias, les truites, le muge (*Mugil cephalus*) et, au Japon, la sériole à cinq bandes (*Seriola quinqueradiata*) [45].

### c) Chez les autres espèces animales

Des souches de *S. agalactiae* sont isolées du singe, du porc, du chien, du chat, de la grenouille, du ragondin, de la souris, du rat, du hamster et du cheval [47-49]. Le pouvoir pathogène est généralement peu documenté et la bactérie est souvent isolée en association avec d'autres bactéries ou des virus. Toutefois, dans quelques cas, l'isolement a été rattaché à un pouvoir pathogène: septicémie chez un singe (*Callithrix jacchus*), endocardite et eczéma chez le chien, syndrome d'affaiblissement chez le chiot, méningo-encéphalomyélites suppuratives chez la souris nude [50, 51].

En conclusion, *S. agalactiae* présente une grande diversité d'hôtes, qui la distingue par exemple de *S. pyogenes*, qui n'a que l'homme comme hôte connu. Au sein de ces différents hôtes, *S. agalactiae* est capable de coloniser différents organes. Comme nous allons le voir, cette adaptabilité à un large éventail d'environnements semble liée d'une part à un polymorphisme génétique entre les souches de *S. agalactiae*, et d'autre part à la richesse en systèmes de régulations de la bactérie.

## II. La diversité des populations

Afin de déterminer si des groupes de souches de *S. agalactiae* sont associés à certains hôtes ou à certains contextes (portage vs infection) différentes méthodes de typage ont été mises en place, ce qui a permis de caractériser la diversité des souches de *S. agalactiae*.

### A. Différentes approches pour le typage de *S. agalactiae*

La première méthode de typage a été basée sur l'analyse de la variabilité structurale de la capsule polysaccharidique de différents isolats de *S. agalactiae* [4]. De nos jours dix sérotypes capsulaires ont été identifiés [52]. Chaque sérotype correspond à un arrangement différent de quatre sucres parmi cinq possibles (glucose, rhamnose, galactose, N-acétylglucosamine et acide acétylneuraminique ou acide sialique). La variété antigénique des polysaccharides capsulaires

constitue la base du système de sérotypage largement utilisé dans l'épidémiologie descriptive de *S. agalactiae*. Puis des méthodes de typage moléculaire lui ont été combinées. La plupart de ces techniques sont basées sur l'analyse de l'ADN comme l'électrophorèse en champ pulsé, l'amplification aléatoire d'ADN polymorphe, les profils de restriction et la distribution d'éléments génétiques mobiles [53-56]. En 2003, un schéma de séquençage de gènes multiples (ou MLST pour Multilocus Sequence Typing) a été proposé [57] et est globalement appliqué au sein de la communauté.

L'approche par MLST repose sur le séquençage de sept gènes de ménage. A chaque séquence unique d'un des sept loci est attribué un numéro d'allèle, leur combinaison donne le profil allélique de la souche appelé "Sequence Type" (ST), et un numéro lui est attribué. La recherche et le dépôt des numéros d'allèles et des STs sont réalisés en ligne sur le site <http://pubmlst.org>. Chez *S. agalactiae* le taux de variabilité observé au niveau des séquences nucléotidiques des 7 gènes s'élève à environ une mutation pour 100 pb. Dans la base de données MLST de *S. agalactiae*, le nombre d'isolats recensés par ST n'est pas mentionné. Cette base de données n'est donc pas un outil épidémiologique, mais elle permet de recenser et de partager la diversité observée pour cette espèce. De nombreuses études utilisant cet outil de typage, ont été publiées, et portent sur la répartition des différentes souches sur différents continents, pour des souches d'origine humaine ou animale.

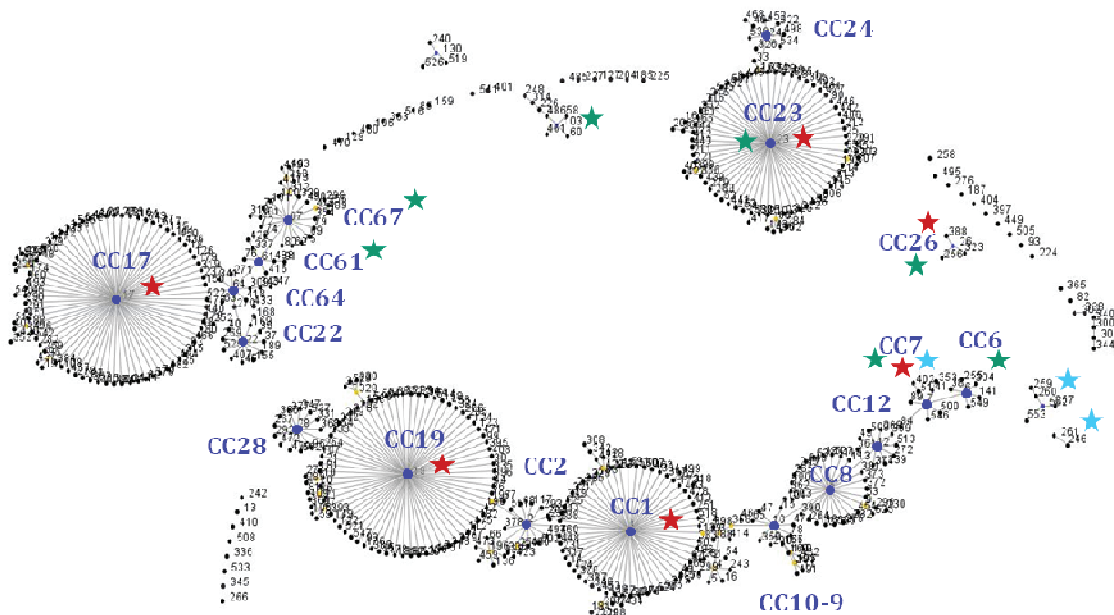
**Tableau 1: Etudes MLST publiées et origines géographiques des souches.**

Continent	Pays	Références
Amérique du nord	Etats-Unis	[58]
	Canada	[59]
Europe	Allemagne	[60]
	Italie	[61]
	Royaume-Uni	[62]
	Suède	[63]
	Portugal	[64]
	France	[11, 29]
	Pologne	[65]
Proche-Orient	Israël	[66]
Afrique	Sénégal - République de centre	[67]
	Afrique	
	Kenya	[68]
Océanie	Australie - Nouvelle Zélande	[69]

L'algorithme eBURST permet de regrouper les isolats présentant un niveau défini de similarité, en complexes clonaux (CC), définis eux même comme les STs partageant au moins

cinq allèles sur les sept [70]. Ce programme détermine ensuite pour chaque CC le clone fondateur. Le clone fondateur est celui qui diffère du plus grand nombre de variants par un seul allèle sur sept. Les SLV (single locus variants) et les DLV (double locus variants) sont regroupés au sein des CCs, et correspondent à des variants qui diffèrent du clone fondateur respectivement par un allèle pour SLV et deux allèles pour DLV. L'image générée par ce programme permet de visualiser la diversité observée pour l'espèce bactérienne étudiée et les relations entre les différents STs et CCs.

## B. Les apports du typage sur l'étude de la diversité



**Figure 3 : Visualisation par eBURST des complexes clonaux reflétant la diversité au sein de la population de *S. agalactiae*.**

La couleur des étoiles indique l'hôte à partir duquel ces souches ont été isolées ; rouge pour les souches humaines, vert pour les souches bovines et bleu clair pour les souches de poissons.

L'analyse par eBURST des 553 STs de GBS permet l'identification de 19 CCs. Parmi ces CCs, certains sont bien définis tel que les CCs 1, 8, 17, 19, 23, 61, et 67. Différentes études analysant le portage de *S. agalactiae* sur les différents continents montrent qu'en Europe et en Amérique du Nord les complexes clonaux CC1, CC17, CC19 et CC23 sont prédominants [11,



29, 57-64]. Ces complexes clonaux sont également retrouvés en Afrique, où un complexe clonal supplémentaire (le CC26) est fréquent et représente 15 % des souches de portage [67].

Différentes études épidémiologiques basées sur le sérotypage ont mises en évidence que les souches de sérotype capsulaire III sont responsables d'une forte proportion des infections précoces et de la plupart des infections tardives. Elles ont aussi mis en évidence l'émergence d'un clone, dit hypervirulent, parmi les souches de sérotype III correspondant aux souches ST17. Ce clone est impliqué dans la plupart des infections néonatales et dans la plupart des méningites. En effet, en France, plus de 80% des souches isolées d'infections tardives sont de sérotype III et la quasi-totalité de ces souches sont du groupe ST17 [11].

De nombreuses études portant sur l'analyse par MLST des souches isolées dans différents hôtes, ont permis de montrer que certains CCs sont spécifiques de l'hôte dans lequel les souches ont été isolées (Figure 3). En effet les CCs 61, 64, 67 et 103 sont spécifiques des souches d'origine bovine [42, 71], les CCs 1, 17 et 19 correspondent à des souches humaines et enfin certains ST correspondent à des souches isolées de poissons tels que les ST260-261 [72]. Cependant d'autres CCs sont retrouvés dans différents hôtes tels que les CC23 et CC26, qui regroupent à la fois des souches bovines et humaines, ou le CC7 qui est constitué de souches humaines, bovines et de poissons [72].

De cette vision sur la diversité des ST, et sur la base des 7 loci du schéma MLST, il a été proposé que les souches ST17 et ST67 ont émergées à partir d'un ancêtre commun bovin [73]. Cependant une analyse comparative récente des souches humaines et bovines, basée sur le séquençage d'un plus grand nombre de gènes, montre que les souches des CCs 17 et 67 forment des groupes indépendants, eux-mêmes éloignés génétiquement des autres complexes [42], et donc que le clone ST17 n'émerge pas d'un ancêtre bovin commun aux souches du CC67. Les souches du CC17 forment un groupe homogène, et possèdent un éventail de facteurs de virulence spécifiques [39, 42, 71]. Enfin, les souches d'origine bovine des CC61 et 67 forment un groupe hétérogène et distant des autres CCs (données du laboratoire sur la base de données génomiques). Par ailleurs, différentes études basées sur l'analyse de combinaisons alléliques de gènes codant pour des protéines de surface, l'hybridation sur puces à ADN et par MLST sur 15 gènes, montrent que le CC23 est constitué en réalité de deux groupes distincts de souches fortement adaptées à leur hôte humain (essentiellement de sérotype Ia) ou bovin (essentiellement de sérotype III) [39, 42].

Les données générées par le MLST permettent d'estimer les contributions relatives entre recombinaisons et mutations dans la diversification de *S. agalactiae*. Pour chaque CC, un SLV dérive de l'ancêtre commun soit par mutation ponctuelle, soit par recombinaison. Un événement de recombinaison est défini par la présence de plus d'une mutation ponctuelle dans l'allèle de différence ou par sa présence dans un autre CC. Le rapport entre les événements de recombinaisons ou de mutations ainsi caractérisé permet de déterminer le taux de recombinaison au sein de l'espèce. Ce ratio par allèle varie de 10:1 en faveur de la recombinaison chez des espèces panmixtiques comme les méningocoques et les pneumocoques (espèces naturellement compétentes) alors qu'il est de 1:15 chez *Staphylococcus aureus*, espèce considérée comme clonale [74]. Chez *S. agalactiae* le ratio recombinaisons/mutations est de 3:1 en faveur de la recombinaison sur l'ensemble de la population [75]. Ce ratio confirme l'importance de la recombinaison homologue et des transferts horizontaux dans la dynamique du génome de *S. agalactiae*. Cependant il existe des différences en fonction des CCs analysés; en effet, pour le CC19 et CC12 le ratio est de 4:1, tandis que pour le CC17 il est de 1:2 confirmant le caractère clonal de ce clone hyper-virulent [76].

### **III. Les facteurs de virulence**

Les infections néonatales à *S. agalactiae* pourraient avoir deux portes d'entrée, une pulmonaire (INP) et une intestinale (INT) [15, 77]. Après avoir traversé ces épithéliums, la bactérie survit et se propage dans le sang. Au niveau du cerveau elle pourrait traverser la barrière hémato-encéphalique déclenchant alors une méningite. La pathogénicité associée à *S. agalactiae* est un processus multifactoriel, qui inclut la capacité de la bactérie à adhérer et à coloniser des épithéliums, puis à envahir les cellules épithéliales et endothéliales et enfin à se multiplier en échappant aux défenses immunitaires de l'hôte [78].

#### **A. *Les différentes étapes de l'infection***

##### **1. Adhésion et colonisation des cellules épithéliales**

L'adhésion aux surfaces épithéliales est la première étape dans le processus de colonisation de l'hôte. *S. agalactiae* colonise différents types d'épithéliums: tels que les épithéliums vaginal, pulmonaire, et vasculaire cérébral. L'adhésion de *S. agalactiae* est favorisée

par un pH acide [79], le pH au niveau du mucus vaginal étant aux alentours de 4,5 [80], cela lui permet de coloniser efficacement cette niche. Il a été constaté que lors des infections, le pH vaginal augmentait allant de 5,5 à 6,5 [80, 81]. Une étude récente a montré que sa survie et sa capacité à produire des biofilm était augmentée avec l'élévation du pH, suggérant que le pH vaginal serait un facteur de risque pour le portage et les infections à *S. agalactiae* [82].

Il a également été démontré que *S. agalactiae* se lie à différentes protéines de la MEC telles que la fibronectine, le fibrinogène et la laminine. *S. agalactiae* se fixe à la fibronectine immobilisée, et non à la fibronectine soluble, qui peut servir d'opsonine pour la reconnaissance par les macrophages [83]. L'adhésion de *S. agalactiae* semble dépendre également de la concentration cytoplasmique de glutamine, en effet des mutants de l'opéron *glnPQ* codant pour le système de transport de la glutamine présentent un défaut de liaison à la fibronectine [84]. Récemment deux protéines (PilA et Srr1 décrites plus loin) connues pour être impliquées dans l'adhésion et la virulence de *S. agalactiae* [85, 86] ont été démontrées comme contribuant à la colonisation vaginale par *S. agalactiae* dans un modèle murin [87]. Différentes structures protéiques à la surface de la bactérie participant au processus d'adhésion à la matrice extracellulaire seront détaillées dans la partie III.B.

## 2. Invasion et traversée des barrières de l'hôte

Au cours du processus infectieux *S. agalactiae* est capable de traverser différentes barrières de l'hôte, telles que la barrière épithéliale pulmonaire ou intestinale et la barrière hémato-encéphalique. La traversée de ces barrières pourrait se faire par invasion et translocation de la bactérie à travers l'épithélium dans une vacuole, dans un processus qui requiert des modifications du cytosquelette d'actine de l'hôte [88], et impliquant la voie de signalisation dépendante de la PI3 kinase [89]. D'autres facteurs pourraient être impliqués dans la pénétration des barrières cellulaires tels que: la hyaluronate lyase qui dégrade le hyaluronane [90], principal composant polysaccharidique des tissus conjonctifs, et le facteur CAMP (initiales du nom des auteurs qui ont décrit le test [91, 92]). Le facteur CAMP est une protéine diffusible qui agit en synergie avec l'hémolysine b de *S. aureus*, et provoque l'hémolyse des hématies, ce qui explique que l'injection, par voie intraveineuse, à des lapins du facteur CAMP soit toxique [93]. Le facteur CAMP s'oligomérisse dans la membrane de la cible et forme des pores conduisant à la lyse cellulaire [94]. Alternativement la traversée peut être liée à une perte d'intégrité de la barrière,

résultant notamment de l'action de la  $\beta$ -haémolysine/ cytolysine ( $\beta$ -H/C) sécrétée par la bactérie [95]. Malgré ces études souvent anciennes, le mécanisme de traversée des barrières par *S. agalactiae* reste encore largement hypothétique.

### 3. Evasion et échappement aux mécanismes de défenses de l'hôte

L'hôte possède deux niveaux de défenses complémentaires pour lutter contre les agressions, et pour limiter l'expansion des bactéries commensales, l'un non spécifique dit inné, et l'autre spécifique dit adaptatif. Le système de défense inné est la première barrière de défense et conduit à la mise en place d'une réponse rapide et non spécifique. Ce système de défense est constitué des barrières physiques telles que la peau et les muqueuses, du complément, et de cellules phagocytaires. Les muqueuses possèdent des glandes qui sécrètent du mucus qui peut piéger et empêcher la progression des bactéries. La muqueuse respiratoire est aussi dotée de cils qui participent à l'élimination de la bactérie. Le pH au niveau de ces muqueuses est également variable, et peut empêcher la prolifération des bactéries. Au niveau des muqueuses sont sécrétées des molécules de défense antimicrobienne [96, 97]. Le plus souvent ces molécules correspondent à des peptides chargés positivement, libérés par les cellules épithéliales et les cellules du système immunitaire. Ces peptides sont attirés par les charges négatives des constituants lipidiques (les phospholipides) des membranes des bactéries. Puis ces peptides s'intègrent dans la membrane et ouvrent des brèches par lesquelles les composés intracellulaires s'échappent [96, 97]. Ces désordres conduisent également à l'activation du système d'autodestruction bactérien (les autolysines). Dans le sang, le complément est un ensemble de protéines agissant par une cascade protéolytique [98, 99]. Ce système agit en participant à la lyse directe du micro-organisme, en facilitant le recrutement et la reconnaissance des bactéries par les cellules phagocytaires, et en initiant la réaction inflammatoire. Puis les cellules phagocytaires jouent un rôle dans l'activation de l'immunité adaptative, et servent de lien entre ces deux systèmes.

*S. agalactiae* exprime différents facteurs sécrétés ou de surface lui permettant d'échapper aux défenses immunitaires innées de l'hôte et participant à sa propagation et à sa survie. Ces différents facteurs agissent au niveau de différents acteurs de la réaction immunitaire. Tout d'abord certains facteurs permettent la résistance aux peptides antimicrobiens, comme la D-alanylation de l'acide lipotéichoïque [100], et les protéines PBP1a (pour Penicilline Binding Protein) [101]. D'autres facteurs interfèrent avec les processus d'opsonisation tels que la C5a peptidase, l'acide sialique de la capsule, la protéine BibA, et la protéine  $\beta$  qui interfèrent avec les

composants du complément C5a, C3, C4 et C3bp. CpsA, une protéase de surface, qui coupe le fibrinogène en fibrine, empêchant le processus d'opsonisation, intervient aussi à ce niveau [102]. Plusieurs facteurs agissent au niveau du processus de phagocytose. Par exemple, la superoxyde dismutase (*sodA*) confère une meilleure résistance au stress oxydatif, qui intervient lors de la destruction par les polynucléaires neutrophiles [103]. Enfin, *S. agalactiae* pourrait échapper aux anticorps de l'hôte en diminuant l'expression d'épitopes antigéniques [104] ou leurs accessibilités [105].

## **B. *Les facteurs de virulence caractérisés chez S. agalactiae***

Les composants de la paroi bactérienne sont situés en première ligne des interactions avec l'hôte, permettant l'adhésion, la dissémination mais aussi la reconnaissance par le système immunitaire, et son échappement. Les éléments de la paroi de *S. agalactiae* jouent un rôle primordial dans l'interaction avec son hôte, et dans la perception et l'adaptation de la bactérie dans son environnement. La majorité des facteurs de virulence identifiés chez *S. agalactiae* sont des composants protéiques ou non, de la paroi [77, 78]. La paroi des bactéries à Gram positif constitue un réseau tridimensionnel complexe composé essentiellement de protéines, de polysaccharides et de polymères anioniques. Chez *S. agalactiae* (Figure 4), la paroi est constituée de différentes structures de type polysaccharidique tels que la capsule, l'acide lipotéichoïque et l'antigène B, mais également de protéines de surface ou de structures protéiques plus complexes tel que les pili.

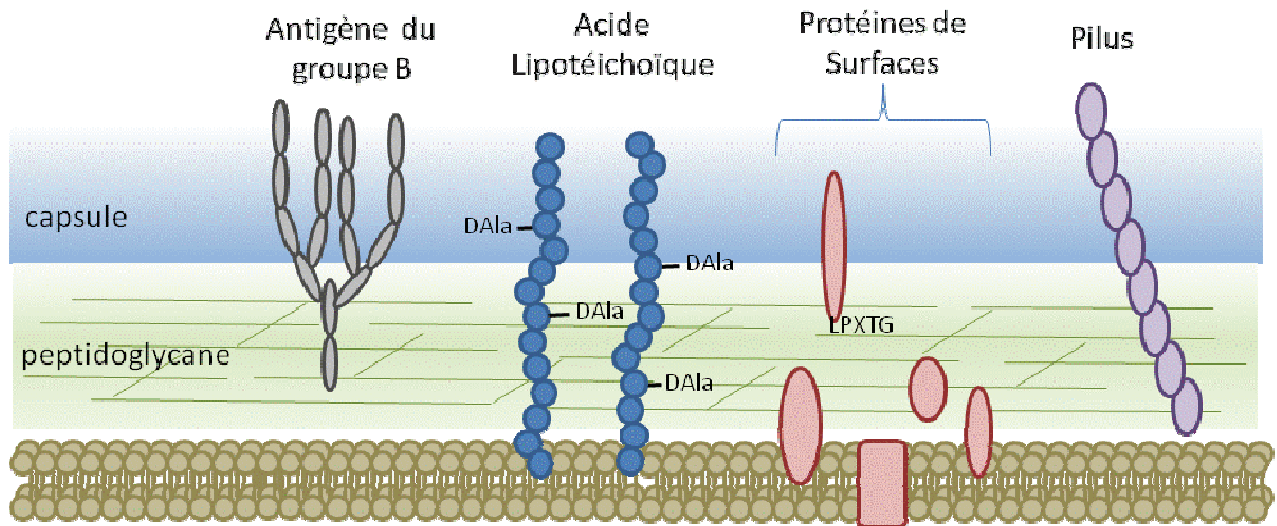
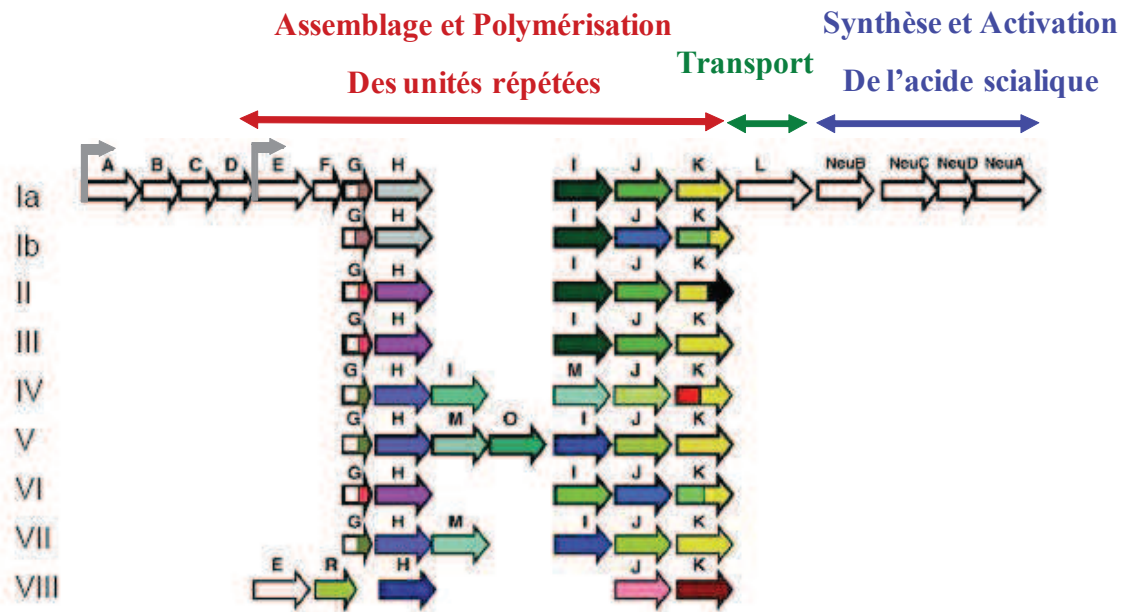


Figure 4 : Paroi de *S. agalactiae*

## 1. Les facteurs de virulence non protéiques

### a) La capsule

La capsule est une structure polysaccharidique localisée à la surface de *S. agalactiae*. En tout, dix sérotypes capsulaires ont été identifiés [52]. Chaque sérotype correspond à un arrangement différent de quatre sucres parmi cinq possibles (glucose, rhamnose, galactose, N-acétylglucosamine et acide acétylneuraminique ou acide sialique). Les enzymes impliquées dans la synthèse de la capsule sont codées au sein d'un même locus, le locus *cps*. L'analyse de ce locus chez huit souches de sérotypes différents montre une organisation similaire, avec des gènes partagés et des gènes spécifiques [106, 107]. Notamment l'expression de glycosyl-transférases de spécificité différentes explique la variabilité dans la combinaison des sucres incorporés [106]. L'étude de différents mutants dans ce locus a permis de montrer que les gènes *cpsE* à *cpsK* sont impliqués dans l'assemblage et la polymérisation des unités répétées, et que *cpsA* code pour un activateur transcriptionnel de la famille LytR [108, 109].



**Figure 5: Comparaison et organisation du locus capsulaire**

Les gènes *cpsA-E*, *cpsL* et *neuA-D* sont conservés dans les 9 sérotypes et sont représentés par des flèches blanches. Les différentes couleurs indiquent les différents allèles des gènes variables. d'après [106]. Le locus du sérotype IX n'est pas indiqué dans ce schéma.

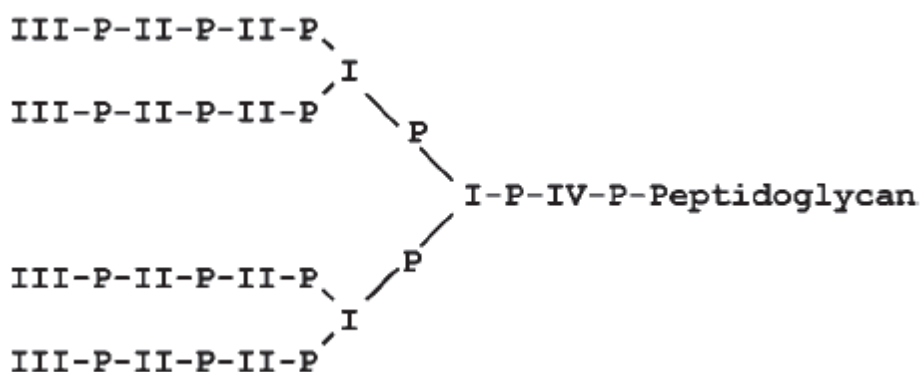
Différentes études ont montré que des souches de *S. agalactiae* sans capsule adhéraient et envahissaient mieux les cellules épithéliales [110, 111] que les organismes encapsulés. Au contraire, la présence de la capsule favorise la survie et l'échappement au système immunitaire de l'hôte. L'exposition à la surface de *S. agalactiae* de molécules sialylées diminue sa fixation au composant C3 du complément par un processus de mimétisme moléculaire [112, 113]. La capsule est un des facteurs de virulence majeur de *S. agalactiae*.

Le contrôle génétique de la production de la capsule est important. Une diminution de la production de polysaccharides capsulaires permet la colonisation et l'invasion, alors que l'augmentation de la synthèse de la capsule favorise la survie de la bactérie dans le sang. La production de capsule serait liée à la vitesse de croissance. En effet, la richesse du milieu de culture et le taux de croissance de la bactérie jouent un rôle important dans la régulation de l'expression de la capsule avec une sous-expression en milieu minimum [114, 115]. Il a également été montré qu'au sein d'un même hôte, des souches isolées de différents tissus présentent une variabilité du degré d'encapsulation [116]. Cette observation suggère la mise en place d'une régulation de l'expression de la capsule de *S. agalactiae* au sein de son hôte. En dehors de *cpsA*, un autre gène codant pour un régulateur de type LysR, *cpsY* (gbs1249), situé en

amont de *cpsA* et transcrit de manière divergente, pourrait être impliqué dans la régulation de la synthèse de la capsule, mais ce rôle reste hypothétique [117].

## b) L'antigène B

L'antigène B est une structure carbohydate localisée à la surface de *S. agalactiae*, à l'origine de son typage en streptocoque du groupe B. Malgré son importance dans le diagnostic, grâce au développement de kits de détection rapide de l'antigène B [118, 119], le locus codant sa synthèse n'a été identifié que récemment [120]. La structure de l'antigène B a été déterminée grâce à différents travaux [121-124]. L'ensemble de ces études a permis de définir une structure multi-antenne basée sur l'arrangement de quatre unités oligosaccharidiques (Figure 6). Chacune de ces quatre unités oligosaccharidiques correspond à un arrangement différent de quatre sucres: le galactose, la N-acétylglucosamine, le glucitol, et le rhamnose [121]. Ces quatre unités oligosaccharidiques sont reliées par des liaisons phosphodiester [124]. L'antigène B est ancré au peptidoglycane au niveau de résidus d'acide N-acétylmuramique. L'analyse de cette structure a permis de définir un minimum de 11 transférases nécessaires à l'élaboration de l'antigène B. A l'exception de trois N-acétylglucosamine transférases, toutes les transférases sont regroupées au sein d'un même locus (*gbs1494-gbs1480*) [120]. Ce locus semble être exprimé en un seul opéron car l'annotation du génome de NEM316 a conduit à l'identification d'un seul terminateur en aval de *gbs1480* [125].



**Figure 6: Structure de l'antigène B**

Les chiffres romains de I à IV correspondent aux quatre unités polysaccharidiques constituant l'antigène B d'après [120]



Peu d'études portent sur l'analyse du rôle de l'antigène B dans la pathogénicité. L'antigène B est un élément proinflammatoire [126]. Des anticorps dirigés contre cet antigène ne semblent pas protecteurs chez la souris [127], peut-être en raison d'une mauvaise accessibilité due à la présence de la capsule [128]. Aucun système de régulation n'a été identifié pour ce composant de la paroi.

### c) L'acide Lipotéichoïque (LTA)

La paroi des bactéries à Gram positif contient communément deux polymères anioniques. Le premier ancré au peptidoglycane est appelé l'acide téichoïque, et le second l'acide lipotéichoïque. L'assemblage des acides téichoïque et lipotéichoïque fait intervenir des voies métaboliques différentes, or *S. agalactiae* ne possède pas certaines des enzymes nécessaires à la synthèse de l'acide téichoïque [129], suggérant qu'il ne possède que le LTA à sa surface. Le LTA est un polymère anionique de polyglycerol-phosphate ancré à la membrane bactérienne par un glycolipide ancreur (Figure 7) [130-132]. Malgré l'importance de ce polymère dans la physiologie bactérienne, très peu de choses sont actuellement connues en ce qui concerne sa synthèse et sa régulation chez *S. agalactiae*.

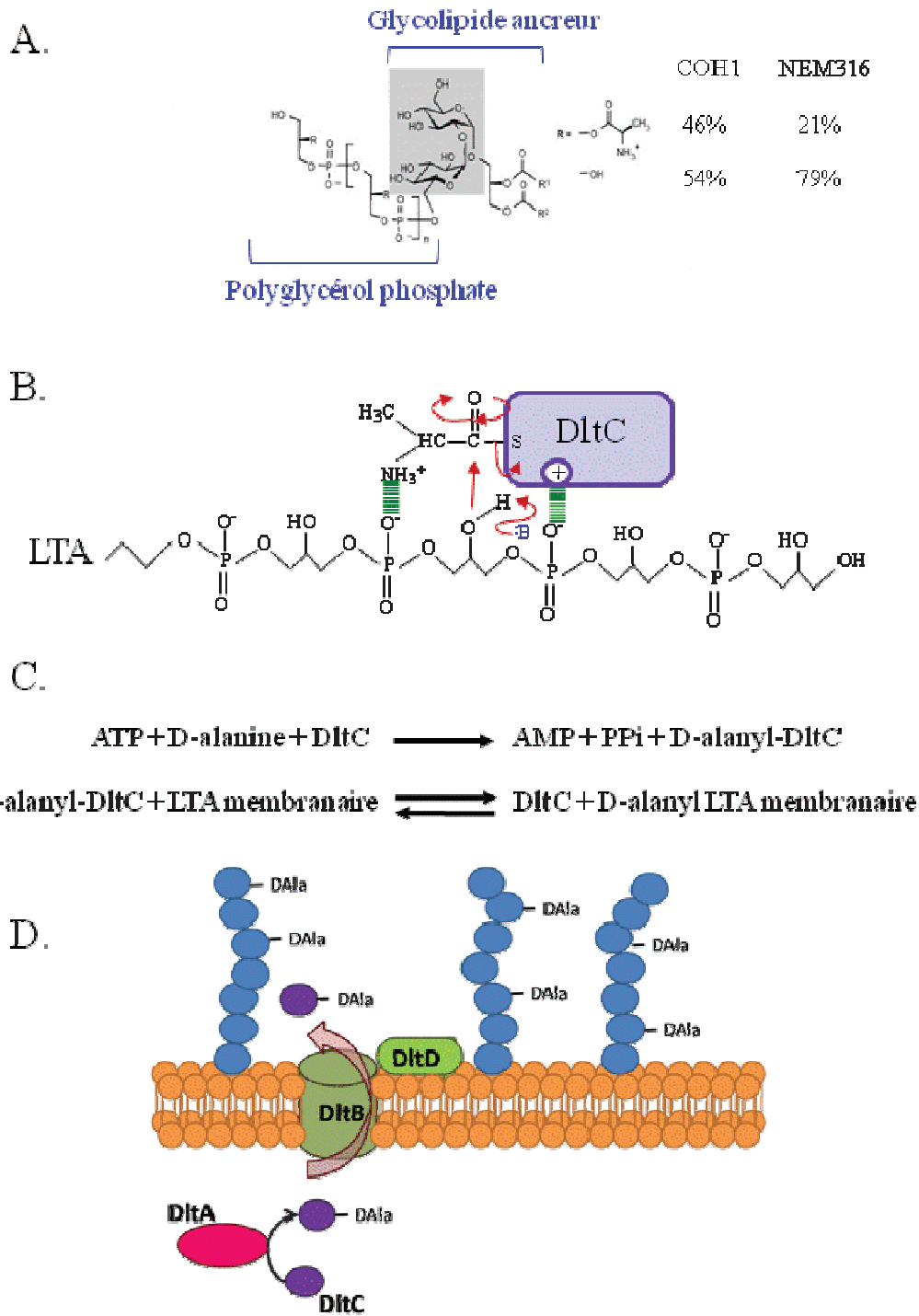
La recherche de mutants atténués dans la capacité à envahir des cellules endothéliales microvasculaires a permis d'isoler un mutant dans le gène *iagA* (gbs0682). Ce mutant relargue du LTA dans le milieu, suggérant que *iagA* est responsable de l'ancrage du LTA à la membrane [133]. Ce mutant n'est pas affecté dans sa capacité à provoquer des septicémies, mais le taux de méningites est réduit avec ce mutant, en comparaison avec la souche sauvage. Un gène codant pour la polyglycérol phosphate polymérase a été caractérisé chez *S. aureus* [134]. Ce gène, *ltaS*, est essentiel pour la croissance à 37°C de *S. aureus* [135]. J'ai pu identifier un homologue de ce gène dans la souche NEM316, *gbs1451*. Mes tentatives pour inactiver ce gène ont été infructueuses, laissant penser que ce gène pourrait être essentiel chez *S. agalactiae* dans les conditions que j'ai utilisées.

Par la présence de ses groupements phosphates ionisés, le LTA confère à la surface de *S. agalactiae* une forte charge négative. Le LTA une fois exposé à la surface de la bactérie peut subir des modifications, telles que l'ajout de résidus de D-alanine sous forme d'esters des groupements hydroxyles de la partie polyglycerol phosphate. La D-alanylation du LTA introduit un groupement amine basique annulant partiellement la charge négative nette. Cette

incorporation de résidus D-alanine au niveau du LTA permet à la bactérie de moduler la charge à la surface de sa paroi. Le taux de D-alanylation du LTA est variable selon les espèces, il est respectivement de 46% et 70% pour *S. agalactiae* dans la souche COH1 et pour *S. aureus* [136]. Le degré de D-alanylation varie en fonction du pH, de la température, de la concentration en NaCl chez *S. aureus* [137-140]. Le processus de D-alanylation repose sur l'activité de 4 protéines (DltA à DltD) codées par les gènes de l'opéron *dlt*, pour "D-alanyl LTA". La protéine DltA code pour l'enzyme responsable du transfert d'un groupement D-alanine à la protéine DltC. Cette protéine est similaire à l'"acyl carrier protein" impliquée dans la synthèse des lipides. DltC chargée de son groupement alanine traverse la membrane bactérienne à travers un tunnel formé par la protéine DltB, constituée de 12 domaines transmembranaires. Puis au niveau de la paroi, le résidu de D-alanyl lié à protéine DltC est transféré au LTA ancré à la membrane (Figure 7.A). Chez *S. agalactiae*, il a été montré que le mutant *dltA*<sup>-</sup> est déficient dans la capacité à incorporer des résidus D-alanine au niveau du LTA [136, 141].

Différentes fonctions ont été proposées pour la D-alanylation: (i) moduler les activités des autolysines chez *S. aureus* et *Lactococcus lactis* [142, 143] (ii) maintenir l'homéostasie et moduler l'assimilation des cations pour les fonctions cellulaires chez *Bacillus licheniformis* [144] (iii) définir les propriétés électromécaniques de l'enveloppe cellulaire.

Selon les espèces, la D-alanylation pourrait avoir des fonctions dans l'adhésion, la formation de biofilms, la sensibilité au pH, la conformation des protéines de surface, la résistance aux peptides antimicrobiens et la virulence [145, 146]. Chez *S. agalactiae*, le mutant *dltA*<sup>-</sup> ne semble pas être altéré de façon significative dans ses propriétés d'adhésion. Cependant, il présente une sensibilité accrue aux peptides antimicrobiens cationiques [100]. Ces peptides antimicrobiens sont présents au niveau des muqueuses et des épithéliums et correspondent à la première barrière de défense de l'hôte; ils agissent en formant des pores au niveau de la membrane bactérienne [147]. L'étude de la virulence du mutant *dltA*<sup>-</sup> dans différents modèles d'infections néonatales chez le rat et la souris, a conduit à la conclusion que la D-alanylation du LTA est un facteur important pour la virulence de *S. agalactiae* [148]. En effet, le mutant *dltA*<sup>-</sup> présente une survie réduite dans le sang et est éliminé plus rapidement des poumons; en outre ce mutant est plus sensible à la phagocytose par les macrophages, dont les macrophages alvéolaires, qui contiennent des protéines antimicrobiennes dans leurs phagosomes. Ces différentes observations permettent d'expliquer que le mutant *dltA*<sup>-</sup> soit incapable d'induire une pneumonie chez la souris [100].



**Figure 7 : Structure et D-alanylation du LTA.**

- A. Formule du LTA et taux de D-alanylation dans les souches NEM316 et COH1 d'après [136, 141].
- B. Mécanisme proposé pour la réaction de transacylation du résidu D-alanyl sur la molécule de LTA d'après [149]
- C. Les étapes de la réaction de D-alanylation du LTA
- D. Localisation des différentes protéines de l'opéron *dlt*

L'opéron *dlt* de *S. agalactiae* est régulé par un TCS *dltRS* situé en amont de *dltABCD* [103]. Le rôle de ce TCS est discuté dans le paragraphe (IV.A.1.b). Jusqu'à récemment, cette organisation était unique à *S. agalactiae*. Des recherches par BLAST sur la banque de données nr du NCBI m'a permis de montrer que cette organisation était conservée dans une espèce éloignée du groupe bovis de *S. agalactiae*: *Streptococcus infantarius* (article 1).

L'incorporation de résidus D-alanine au niveau du LTA joue un rôle majeur dans la capacité de colonisation de *S. agalactiae*, car elle permet à GBS de moduler la charge à sa surface, ce qui modifie notamment sa sensibilité aux peptides cationiques antimicrobiens. La première partie de ma thèse porte sur la régulation de l'expression de l'opéron *dlt* et l'impact de l'absence de D-alanylation sur le profil global de transcription chez *S. agalactiae*.

## 2. Les structures protéiques

### a) Les protéines Alpha like et Beta

Historiquement l'antigène C fut un des premiers antigène de surface identifié chez *S. agalactiae*. Il a été montré que cet antigène correspondait à deux protéines indépendantes: la protéine  $\alpha$  résistante à la trypsine et la protéine  $\beta$  sensible à la trypsine. Cinq protéines appartenant à la famille Alpha like (Alp) et caractérisées par une structure hautement répétitive ont été identifiées: Rib,  $\alpha$ , Alp2, R28 (ou Alp3) et Epsilon [39, 40]. La quasi-totalité des souches de *S. agalactiae* exprime un seul des cinq membre de cette famille. La protéine  $\alpha$  située à la surface de *S. agalactiae*, confère une immunité protectrice chez la souris, et les réactions croisées sont limitées. Elle interagit avec le glycosaminoglycane (GAG) de l'hôte par son domaine N-terminal [150], et participe à l'entrée et à la traversée des cellules épithéliales, par un mécanisme qui impliquerait un réarrangement du squelette d'actine [88]. La protéine Rib, apparentée à la protéine R28 de *S. pyogenes*, favorise la fixation de la bactérie aux cellules épithéliales [151].

La protéine  $\beta$  (Bac) codée par le gène *bca* situé dans un transposon composite confère également une immunité protectrice chez la souris. Il a été montré que la protéine Bac se lie aux sites Fc des IgA et du facteur H (FH). La capacité de Bac à se lier aux IgA permet à la souche A909 échapper à la réponse immunitaire, normalement déclenchée par la liaison des IgA au CD89 [152]. De plus FH régule la voie alternative du complément, et FH lié à Bac conserve sa

capacité à réguler négativement l'activation du complément [152]. L'importance de la protéine  $\beta$  dans la virulence semble être restreinte aux souches du complexe clonal 7; en effet parmi les souches séquencées, le gène *bac* n'est retrouvé que dans les souches A909 et H36B. Ce gène est sous la forme d'un pseudogène dans les autres souches séquencées. Une étude récente a montré que la transcription du gène *bac* est contrôlée par le régulateur BrgR [153] appartenant à un système à deux composants (Sak188/Sak189), localisé en amont du gène *bac* dans le même îlot génomique, de la souche A909.

## b) La Laminine Binding Protein (Lmb) et la C5a peptidase

La protéine Lmb est une lipoprotéine qui permet l'attachement de *S. agalactiae* à la laminine humaine. La laminine est un des constituants majeurs de la lame basale et de la MEC. La protéine Lmb participe à l'adhésion et à la colonisation des épithéliums [154]. Elle serait aussi impliquée dans l'invasion des cellules endothéliales microvasculaires [155]. Le gène *lmb* est situé dans un transposon composite de 16 kb flanqué de deux séquences d'insertion (IS) ISSag2 [156]. Ce transposon code également pour une autre protéine de surface la C5a peptidase (gène *scpB*) qui est ancrée au peptidoglycane par un motif LPXTG. La C5a peptidase est, elle aussi, importante pour l'interaction de la bactérie avec son hôte. Elle est constituée de deux domaines. Le domaine protéase N-terminal est impliqué dans l'inhibition de l'activation du système immunitaire par clivage du composant C5a du complément. Le domaine C-terminal correspond à un domaine d'interactions à la fibronectine [157, 158]. Deux études montrent que la plupart des souches d'origine humaine portent ces deux gènes, alors que beaucoup de souches animales ne les possèdent pas [39, 156]. Il semblerait donc qu'il y ait une pression de sélection chez l'homme pour la conservation de ces deux protéines, suggérant que ces protéines pourraient être importantes pour le portage chez l'homme, mais pas chez l'animal. Les gènes *lmb* et *scpB* sont pratiquement identiques (99 et 98% d'identité) aux gènes *lmb* et *scpB* trouvés dans le "core" génome de *S. pyogenes*. Il est donc probable que ces gènes furent acquis par un transfert horizontal récent.

La région intergénique entre *scpB* et *lmb* est un point chaud d'intégration pour l'IS1548 et l'intron du groupe II GBSi1. L'IS1548 est majoritairement retrouvée dans les souches du CC19.

GBSi1 est lui principalement trouvé dans les souches du CC17. Les souches possédant l'IS1548 présentent une expression accrue du gène *lmb*, conduisant à une augmentation de la quantité de protéine Lmb à la surface, conférant une liaison accrue à la laminine. La délétion de l'IS dans ces souches montre que ce phénotype est bien lié à la présence de l'IS [159]. Aucun régulateur de *S. agalactiae* n'a été identifié comme régulant directement l'expression de ces deux protéines. En particulier l'expression de ces deux gènes n'est pas modifiée dans le mutant  $\Delta covRS$ .

### c) Les protéines FbsA et FbsB

Les gènes *fbsA* (*gbs1087*) et *fbsB* (*gbs0850*) codent pour des protéines impliquées dans la liaison au fibrinogène. La protéine FbsA est une protéine ancrée à la surface de la bactérie par son motif LPXTG. Elle comporte un motif de 16 acides aminés impliqués dans la liaison au fibrinogène [160] et répété de 3 à 30 fois selon les souches de *S. agalactiae*. La seconde protéine de liaison au fibrinogène est FbsB/Fgag. Elle a été identifiée de façon indépendante dans une souche humaine [161] et dans une souche bovine [162]. La protéine FbsB, ne possède pas de motif LPXTG, ce qui suggère, qu'elle n'est pas ancrée à la surface, mais est sécrétée dans le milieu extracellulaire [162]. De plus, il a été montré que la protéine de la souche bovine se lie au fibrinogène bovin mais pas humain [162], ce qui confirme l'adaptation de cette protéine à l'hôte bovin.

Les souches hypervirulentes du complexe clonal CC17 et de sérotype III présentent une capacité accrue de liaison au fibrinogène [163], dans laquelle la protéine FbsB joue un rôle majeur. Ce résultat diffère de ceux obtenus pour la souche du CC23 et de serotype III (souche 6313), chez laquelle le rôle majeur est joué par FbsA [160]. De manière intéressante, ces souches des CC 17 et 23 possèdent les deux gènes *fbsA* et *fbsB*, mais les régulateurs de ces gènes sont différenciellement présents. En effet contrairement aux souches du CC17, la souche 6313 possède un locus *rgfBDAC* non fonctionnel et code pour le régulateur *RogB* absent des souches du CC17 [164]. (CF paragraphe sur les systèmes de régulation)

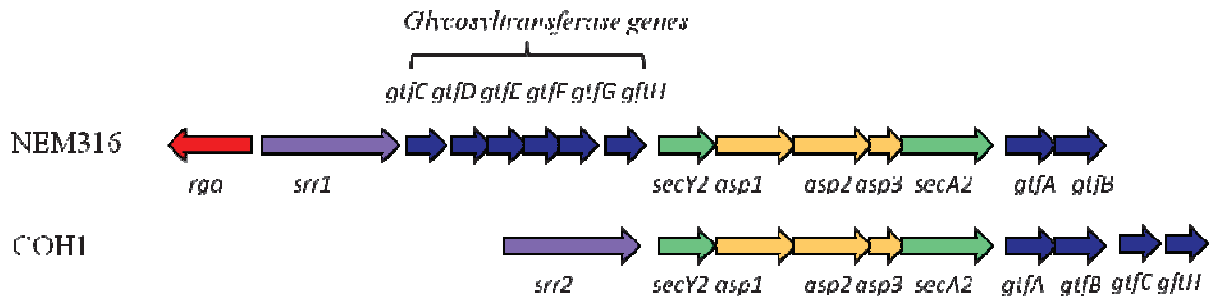
### d) La protéine riche en sérine Srr du locus *secA2*

Des glycoprotéines riches en sérine (constituée sur 75% de la protéine de répétition riches en sérines SX) ont été identifiées dans différentes espèces de streptocoques, staphylocoques et

lactobaciles [165]. Ces protéines sont ancrées au peptidoglycane par un motif LPXTG. Les protéines de cette famille ont des fonctions diverses dans l'interaction avec l'hôte. La protéine Hsa de *S. gordonii* permet l'adhésion à la fibronectine humaine, et à des glycoprotéines membranaires de l'hôte [166-168]. Le gène codant pour cette protéine, *srr*, est inclus dans un locus appelé locus *secA2*, codant un système de sécrétion alternatif. Le locus *secA2* de la souche NEM316 est constitué (i) des gènes impliqués dans la formation du système de sécrétion: *secA2*, *secY2* et au moins trois gènes codant pour des protéines accessoires, (ii) de sept glycosyltransférases (iii) d'une protéine sécrétée riche en sérine, Srr1. Une analyse comparative de la diversité des souches de *S. agalactiae* a indiqué la présence de deux types de locus *secA2* structurellement apparentés à ceux trouvés chez d'autres streptocoques. Le premier locus est retrouvé au sein de souches appartenant à différents ST et code pour la protéine Srr1, alors que le second locus semble restreint aux souches du CC17 hypervirulentes [39].

La protéine Srr1 est ancrée à la surface de *S. agalactiae* par son motif LPXTG. Après synthèse, la protéine est d'abord glycosylée, puis sécrétée par le système SecA2 et ancrée à la surface de la bactérie via la sortase A. Les glycosyltransférases GtfA et GtfB, responsables de l'attachement de résidus de N-acétylglucosamide (GlcNac) et d'acide sialique, sont essentielles pour l'export de la protéine Srr1 à la surface de la bactérie. Ce résultat indique que Srr1 pourrait participer à l'évasion de *S. agalactiae* aux défenses de l'hôte, en contribuant à la présentation d'acide sialique à la surface de la bactérie. Par ailleurs, la glycosylation de Srr1 rend la protéine plus résistante aux attaques protéolytiques. Il a été montré que Srr1 se lie à la kératine humaine K4, grâce à un domaine de 157 AA situé en C-terminal, et participe l'adhésion de *S. agalactiae* aux cellules hépatiques HEP-2 [169]. Srr1 contribue aussi à l'adhésion de la bactérie aux cellules épithéliales humaines et à la virulence de la bactérie dans des modèles d'infections néonatales de rat. La glycosylation de Srr1 par les glycosyltransférases GtfC à GtfH permet la modulation de l'adhérence bactérienne et de la virulence [86].

L'expression de *srr1* est sous le contrôle du régulateur Rga codé par le gène divergent situé en amont de *srr1*. Cependant le régulateur Rga ne contrôle pas l'expression des gènes en aval de *srr1*, impliqués dans le transport et les modifications post-traductionnelles de la protéine. Une analyse transcriptomique récente montre que l'expression de *srr1* est induite lors d'une incubation dans le sang [170].



**Figure 8: Organisation du locus *secA2* dans les souches NEM316 et COH1 de *S. agalactiae*.**

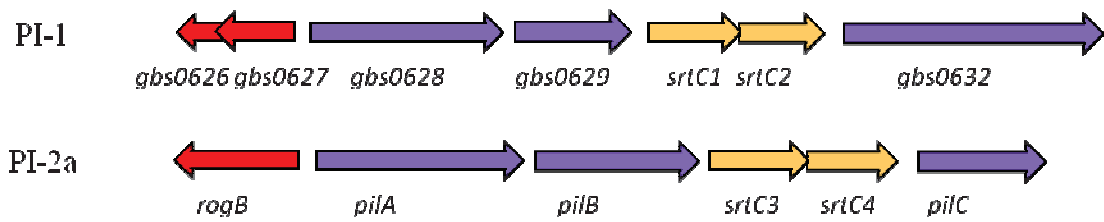
Les gènes codant pour le régulateur transcriptionnel Rga, la protéine LPXTG Srr, les glycosyltransférases sont respectivement colorés en rouge, violet, et bleu. Les gènes correspondant aux protéines accessoires sont en beige et ceux codant pour les deux protéines du système de sécrétion *secY2* et *secA2* sont en vert. (d'après [86])

### e) Les pili

*S. agalactiae* expose à sa surface des structures de type pilus. Ces structures sont constituées de 3 protéines à motif LPXTG ou pilines: une piline majeure, et deux pilines auxiliaires [85]. La première unité auxiliaire est ancrée à la sous-unité principale. La deuxième permet l'ancrage du pilus au peptidoglycane. L'assemblage du pilus est sous le contrôle de la sortase de type A, codée par le "core" génome, qui permet l'ancrage au peptidoglycane [171-173], et de deux sortases de type C qui catalysent la polymérisation des pilines.

Les gènes codant pour les trois pilines et les deux sortases C sont regroupés dans trois types d'îlots génomiques localisés dans deux loci (PI-1 et PI-2), le locus PI-2 présentant deux variants PI-2a et PI-2b. En supplément, l'îlot PI-1 code pour un régulateur transcriptionnel de type AraC (frameshifté dans la souche NEM316) et une protéine de choc thermique, alors que l'îlot PI-2a code pour un régulateur additionnel de type RogB, lequel est absent dans l'îlot PI-2b dépourvu de gènes régulateurs.





**Figure 9 : Loci pili chez NEM316**

Les gènes codant pour les régulateurs transcriptionnels et les protéines LPXTG sont respectivement indiqués en rouge et en violet. Les gènes codant pour les quatre sortases de type C sont de couleur beige. Dans le locus PI-2a constitué des gènes *gbs1479* à *gbs1474* chez NEM316, *gbs1478* (PilA), *gbs1477* (PilB), et *gbs1474* (PilC) codent pour les pilines, et *gbs1476-75* codent pour les deux sortases C (SrtC3 et SrtC4). PilB, correspond à la piline majeure; PilA, à l'adhésine associée au pilus, et PilC, à une protéine accessoire [85].

Différentes études ont montré que les protéines PilA et PilB du pilus PI-2a sont impliquées dans l'adhésion à différents types cellulaires, dont les cellules endothéliales de la barrière hématoencéphalique [174], les cellules pulmonaires [85], et contribueraient à la translocation à travers les barrières épithéliales [175]. PilA et PilB de PI-2a participent également à la formation de biofilms [173, 176]. Récemment, il a été démontré que la sous-unité PilB est un facteur de virulence important dans un contexte néonatal chez la souris, pour la souche NEM316, mais ne semble pas être impliquée, dans cette souche, dans les mécanismes de résistance aux défenses immunitaires innées de l'hôte; en effet, aucune différence n'a été observée entre le mutant  $\Delta pilB$  et le sauvage pour la phagocytose et la survie dans les macrophages, ni pour la résistance aux peptides antimicrobiens (AMP-1D, LL-37, colistin, et polymyxin B) [177]. Ces résultats sont en conflit avec ceux observés dans la souche hyper-hémolytique NCTC10/84 de sérotype V [178] dans laquelle PilB participe à la résistance aux macrophages et aux neutrophiles et à la résistance aux peptides antimicrobiens. Cette dernière étude montre également que la surexpression du gène *pilB* chez *L. lactis* augmente la résistance de cette bactérie à la phagocytose, et rend cette souche virulente dans un modèle murin.

L'expression du gène *pilA* du locus pilus PI-2a est sous le contrôle du régulateur RogB (Gbs1479) [85] et sous le contrôle du régulateur Rga, également de la famille RofA-like, associé au locus *srr-secA2* [179].

#### f) L'adhésine Gbs2018

Cette protéine de surface est ancrée au peptidoglycane par son motif LPXTG. Les analyses comparatives des génomes de *S. agalactiae* [39, 180] ont mis en évidence la structure mosaïque

du gène codant pour cette famille de protéines. Dans les différentes souches, la région promotrice est conservée, ainsi que les régions du gène codant pour les extrémités N et C terminales de la protéine, mais les parties internes sont différentes. Les deux variants majoritaires sont Gbs2018A, également appelé BibA [181], et Gbs2018C, qui est strictement spécifique des clones hypervirulents ST17 [182], et est appelé HvgA pour « Hypervirulent GBS adhesin » [15].

BibA est impliqué dans la survie de GBS dans le sang humain et la résistance à la phagocytose par les neutrophiles humains, et se lie à la protéine C4 humaine du complément [181]. Des expériences d'adhésion à différents types cellulaires, de deux souches de référence BM110 (sérotipe III et ST17) et NEM316 (sérotipe III et ST23), qui expriment respectivement HvgA et un variant de BibA, montrent que ces deux souches adhèrent de façon similaire aux cellules épithéliales pulmonaires. Cependant la souche BM110 adhère plus efficacement aux cellules épithéliales intestinales, aux cellules de la barrière hémato-encéphalique et aux cellules endothéliales. Le remplacement de BibA par HvgA dans la souche NEM316 confère à cette souche un phénotype d'adhésion similaire à celui des souches ST17 [15]. En outre cette étude montre également comment un facteur impliqué dans la colonisation intestinale peut devenir un facteur de virulence dans un contexte néonatal, dans lequel la flore intestinale n'est pas encore établie et ne joue pas son rôle protecteur vis-à-vis des bactéries potentiellement pathogènes [15]. Comme cela a été démontré pour Gbs2018A/BibA [183] la transcription du gène *hvgA* est activée dans un mutant  $\Delta CovR$ .

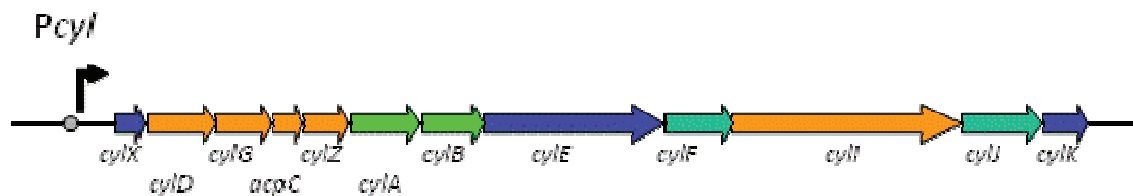
### 3. L'opéron *cyl*

*S. agalactiae* est une bactérie bêta hémolytique, ce qui signifie que, sur gélose au sang de mouton, cette bactérie est capable de détruire complètement les globules rouges, conduisant à l'apparition d'une zone claire autour de la colonie. L'hémolyse bêta fait partie des premiers critères recherchés lors du dépistage de *S. agalactiae*. L'opéron *cyl* a été identifié par la recherche; au sein d'une banque, des mutants ayant perdu cette capacité hémolytique [184]. L'opéron *cyl* est constitué de 12 gènes *cylX*, *cylD*, *cylG*, *acpC*, *cylZ*, *cylA*, *cylB*, *cylE*, *cylF*, *cylI*, *cylJ*, *cylK*, et est localisé dans l'îlot VI de NEM316. Ces gènes codent pour: un transporteur ABC (*cylA*, *cylB*), une protéine homologue à l'acyl carrier protéine (*acpC*), quatre protéines potentiellement impliquées dans la biosynthèse des acides gras (*cylD*, *cylG*, *cylZ*, *cylI*), et une protéine agissant potentiellement comme glycosyltransférase (*cylJ*) apparentée aux UDP-

glucuronosyltransferases [185]. Jusqu'à aujourd'hui 9 des 12 gènes de l'opéron *cyl* ont été inactivés (*acpC*, *cylZ*, *cylA*, *cylB*, *cylE*, *cylF*, *cylI*, *cylJ*, et *cylK*). Les mutants sont tous hémolytiques, à l'exception du mutant *cylE* [95], indiquant que la protéine CylE serait seul responsable de l'activité hémolytique de *S. agalactiae* et de la pigmentation associée. Néanmoins la nature de cette hémolysine reste énigmatique. Elle n'a pas par exemple été purifiée, et le lien avec les autres gènes de l'opéron reste inconnu.

La  $\beta$  hémolysine/cytolysine ( $\beta$ -H/C) participe à l'échappement aux défenses immunitaires par son action cytolytique sur les polynucléaires neutrophiles et les macrophages [186]. L'action de cette toxine conduit à la formation de pores, entraînant la lyse des cellules épithéliales pulmonaires et des cellules endothéliales. Un mutant de la  $\beta$ -H/C présente une diminution de sa capacité à pénétrer la barrière hémato-encéphalique dans un modèle murin [187]. Le pigment orangé intervient également dans la survie à l'intérieur des phagocytes en raison de ses propriétés antioxydantes [186, 188]. Ces deux facteurs agissent conjointement pour lutter contre les défenses immunitaires phagocytaires [186].

L'expression de l'opéron *cyl* est reprimée par le système à deux composants CovRS, et est activée par le régulateur simple RovS [183, 189, 190].



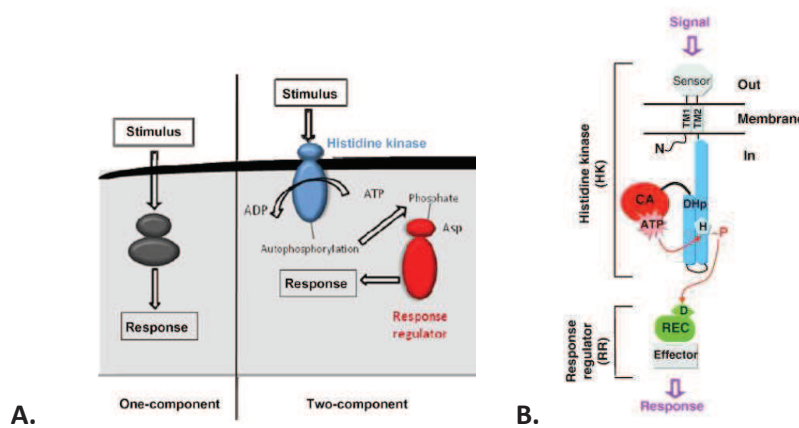
**Figure 10: Organisation de l'opéron *cyl* chez *S. agalactiae*.**

Les gènes codant pour l'ABC transporteur sont indiqués en vert, les gènes codant pour des protéines potentiellement impliquées dans la synthèse des acides gras sont indiqués en orange, les autres gènes de l'opéron sont en bleu. La présence d'un site de fixation de CovR en amont du promoteur *P<sub>cyl</sub>* est indiquée par la présence d'un rond gris.

#### **IV. *S. agalactiae* une bactérie riche en systèmes de régulation**

Les données de l'annotation du génome de la souche de *S. agalactiae* NEM316 ont permis d'identifier 107 gènes codant pour des régulateurs transcriptionnels potentiels, représentant 5% des gènes prédits. En ce qui concerne les systèmes de transduction du signal utilisés chez les procaryotes, la plupart impliquent des systèmes à deux composants (TCSs). Ces systèmes sont constitués d'une protéine histidine kinase (HK) située le plus souvent au niveau de

la membrane et d'une protéine régulatrice (RR, pour « response regulator ») (Figure 11). La protéine HK est constituée de quatre domaines: le domaine senseur qui est impliqué dans la perception du signal, un domaine constitué de deux hélices transmembranaires, le domaine catalytique (CA) également impliqué dans la liaison à l'ATP, et le domaine de dimérisation (DHP) qui contient le résidu histidine conservé qui va être phosphorylée lors de l'activation du système. L'autophosphorylation au niveau d'un résidu histidine de la protéine HK est régulée par des stimuli environnementaux, puis le groupement phosphate est transféré à un résidu aspartate de la protéine régulatrice, qui agit comme régulateur transcriptionnel, conduisant à l'activation ou la repression des gènes cibles.



**Figure 11: Les systèmes à deux composants et voies de transduction du signal**

**A.** Différences entre les voies de transduction du signal impliquant soit des régulateurs simples, soit des régulateurs appartenant à des systèmes à deux composants (d'après [191]). **B.** Représentation des domaines de l'Histidine Kinase et du régulateur transcriptionnel impliqués dans la voie de signalisation des systèmes à deux composants (d'après [192]).

De nombreuses bactéries pathogènes sont capables de s'adapter aux différents environnements qu'elles rencontrent au cours de leur cycle d'infection, cette capacité d'adaptation faisant souvent intervenir des TCSs. Le génome de la souche de *S. agalactiae* NEM316 code pour 20 systèmes à deux composants. Ce nombre est plus élevé que dans les espèces proches: 14 chez *S. pneumoniae* et *S. mutans*, 13 chez *S. pyogenes*, 10 pour *S. thermophilus*, et seulement 8 chez *L. lactis* [193-196]. Cette observation suggère que *S. agalactiae* doit répondre plus finement à différents paramètres environnementaux ou à un plus grand nombre de stimuli, comparé à ces espèces phylogénétiquement proches. Le système de

transduction du signal chez *S. agalactiae* est également constitué d'une Sérine/Thréonine kinase (Stk1) [197], et sa phosphatase antagoniste Stp1.

**Tableau 2: Liste des régulateurs transcriptionnels de la souche NEM316 de *S. agalactiae***

<p>87 Régulateurs simples 21 familles</p>	<p><i>hrcA</i> (gbs0094), 6 famille LacI (gbs0118, gbs0804, gbs1509, gbs1736, gbs1849, gbs1884), 4 famille MarR (gbs0150, gbs0330, gbs0816, gbs1402), 13 famille Cro/cI/XRE (gbs0168, gbs0207, gbs0212, gbs0475, gbs0833, gbs1147, gbs1196, gbs1327, gbs1344, gbs1761, gbs1972, gbs1994, gbs2117), 1 famille BglG (gbs0191), 2 famille rgg/GadR/MutR (gbs0230, gbs1555), 6 famille TetR/AcrR (gbs0250, gbs0468, gbs0617-18, gbs1381, gbs1693, gbs2094), 6 famille DeoR/ SorC (gbs0314, gbs0315, gbs1337, gbs1418, gbs1923, gbs1975), 2 famille LytR (gbs0355, gbs0612), 1 famille CopY (gbs0420), 1 famille Fur (gbs0427), 5 famille MerR (gbs0462, gbs1699, gbs1793, gbs1807 (glnR), gbs1958), 5 famille AraC/XylS (gbs0458, gbs0469, gbs0626-27, gbs1173, gbs1221), 3 ArgR (gbs0546, gbs2055, gbs2119), 4 famille LysR (gbs0666, gbs1249 (cpsY), gbs1437, gbs1582), <i>ccpA</i> (gbs0680), 1 famille OmpR (gbs0685), 5 famille GntR (gbs0672, gbs0928, gbs0954, gbs1201, gbs2027), 1 famille CoiA (gbs0823), BirA (gbs0848), 2 famille PhoU (gbs1022, gbs1949), 3 famille RofA (gbs1426, gbs1479, gbs1530), 2 famille PyrR et PurR (gbs1434, gbs1815), 1 famille CodY (gbs1719), 2 famille Fe_dep_rprC (gbs1590, gbs1749), 3 famille FNR/CRP (gbs1780, gbs1882, gbs2120), CtsR (gbs1870), famille inconnue (gbs0267, gbs0342).</p>
<p>20 systèmes à deux composants</p>	<p><i>gbs0121-22</i>, <i>gbs0180-81</i>, <i>gbs0298-99</i>, <i>gbs0309-10</i> (LiaSR), <i>gbs0429-30</i>, <i>gbs0597-98</i> (VncRS), <i>gbs0741-42</i> (walRK), <i>gbs0963-64</i> (graRS), <i>gbs1019-1020</i> (ciaRH), <i>gbs1051-52</i>, <i>gbs1397-98</i> (relRS), <i>gbs1671-72</i> (covRS), <i>gbs1834-35</i> (dltRS), <i>gbs1908-09</i> (sivSR), <i>gbs1934-35</i>, <i>gbs1943-44</i> (rgfCA), <i>gbs1947-48</i> (phoRB), <i>gbs2009-10</i>, <i>gbs2081-82</i>, <i>gbs2086-87</i>.</p>

Au moment de la rédaction de cette introduction peu de ces systèmes de régulations ont été étudiés en détails. La plupart des régulateurs étudiés jusqu'ici correspondent à des régulateurs localisés à proximité de gènes codant pour des protéines de surface ou des facteurs de virulence, ou sont homologues à des régulateurs intervenant sur l'expression de protéines de surface et de facteurs de virulence d'autres streptocoques, en particulier *S. pyogenes*.

## A. Les systèmes à deux composants

### 1. CovRS

Le système CovRS, également appelé CsrRS, a été identifié à l'origine chez *S. pyogenes*, où sa mutation entraîne un phénotype de colonies mucoïdes, en raison de l'augmentation de la production de la capsule d'acide hyaluronique [198]. Le mutant  $\Delta covR$  de *S. pyogenes* est hyper virulent dans un modèle d'infections invasives chez la souris et présente une résistance accrue à

la phagocytose par des leucocytes humains [198-200]. Par ailleurs des mutants des gènes *covR* et *covS* sont fréquemment isolés d'infections invasives [201, 202]. Des données transcriptomiques montrent que CovR réprime directement ou indirectement 15% des gènes de *S. pyogenes*, dont de nombreux gènes codant pour des protéines de surface ou des protéines sécrétées, qui sont potentiellement impliquées dans les interactions entre la bactérie et son hôte [203].

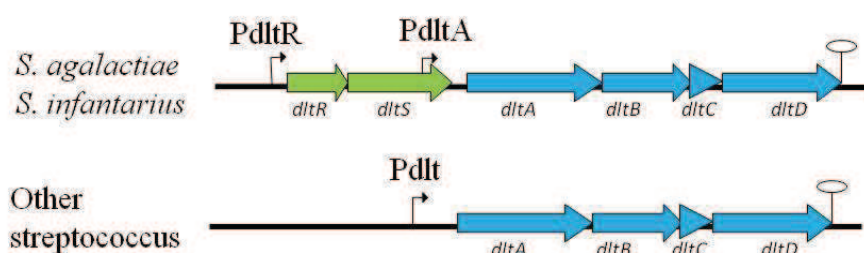
Chez *S. agalactiae*, le mutant  $\Delta covRS$  présente de nombreux changements phénotypiques: il est incapable de se multiplier en milieu minimum RPMI, et sa viabilité dans du sérum humain est fortement affectée en comparaison avec la souche sauvage. En outre, ce mutant présente une activité hémolytique accrue et son activité CAMP est fortement diminuée. Ce mutant présente également une meilleure adhésion aux cellules épithéliales et une virulence accrue dans des modèles d'infections néonatales chez le rat. Les données de transcriptomes indiquent que, chez *S. agalactiae*, tout comme chez *S. pyogenes*, CovRS joue un rôle central. L'expression de 76 gènes est induite et celle de 63 gènes est réprimée dans le double mutant, soit 7% des gènes du génome de GBS [183]. De nombreux gènes codant pour des composants de surface sont induits dans le double mutant (*fsbA*, *gbs0451*, *gbs0791*, *gbs1087*, *bibA*), ainsi que l'opéron *cyl*. Au contraire, l'opéron codant pour la capsule semble un peu réprimé chez le mutant, permettant aux adhésines de surface d'être potentiellement plus accessibles. Le gène *cfb* (*gbs2000*) codant pour le facteur CAMP est lui aussi réprimé. Enfin les gènes codant pour trois systèmes de type ABC sont fortement réprimés, les défauts de croissance observés pouvant être liés à ces modifications métaboliques.

Chez *S. agalactiae*, la fixation spécifique de CovR aux régions promotrices des gènes *gbs2018*, *gbs1895*, et *cylX* a été montrée *in vitro*. L'analyse comparée de ces régions d'ADN a permis d'identifier une séquence de 9 nucléotides conservés 5'TATTTAAT3' [183]. Ce motif semble nécessaire, mais pas suffisant pour la reconnaissance des promoteurs cibles de CovR. Il a été montré que la phosphorylation de CovR par CovS, au niveau du résidu aspartate 53, stimulait sa liaison aux séquences promotrices cibles chez *S. pyogenes* [204] et chez *S. agalactiae* [197], augmentant respectivement l'effet répresseur et l'effet activateur de CovR sur *cylE* et sur *cfb* [189]. Au contraire, une autre étude réalisée chez *S. agalactiae* indique, que la Serine/Thréonine Kinase Stk1 induit l'expression de *cylE*, et réprime celle du gène *cfb*, et que son action est dépendante de CovR [190]. Il a été démontré que la phosphorylation de CovR par Stk1 au niveau de la thréonine 65 le rend inactif [197].

Bien qu'il soit reconnu que, chez *S. pyogenes*, le senseur CovS répond à des modifications de la concentration extracellulaire de  $Mg^{2+}$  [205, 206], chez *S. agalactiae* cela ne semble pas être le cas et le stimulus ressenti par CovS n'est pas identifié [189]. Cette différence peut être liée à la faible conservation du domaine N-terminal de la protéine senseur CovS chez ces deux espèces (36% d'identité).

En résumé, bien que les systèmes orthologues CovRS chez *S. agalactiae* et *S. pyogenes* contrôlent des gènes appartenant aux mêmes classes fonctionnelles (facteurs de virulence, protéines de surface ou sécrétées), les gènes régulés par ces systèmes ne sont généralement pas orthologues, alors que les deux espèces sont proches phylogénétiquement. Ces systèmes semblent également ne pas répondre aux mêmes stimuli dans les deux espèces. Ces différences reflètent probablement l'adaptation des deux pathogènes aux environnements spécifiques qu'ils rencontrent. En outre, ces résultats montrent l'existence de deux systèmes de phosphorylation impliqués dans le contrôle de l'activité d'un même régulateur transcriptionnel, permettant ainsi une régulation fine de l'expression de nombreux facteurs importants pour l'interaction avec l'hôte.

## 2. DltRS



**Figure 12 : Organisation de l'opéron *dlt* au sein du genre streptocoque d'après [141]**

Contrairement aux autres bactéries à Gram positif, *S. agalactiae* possède deux gènes *dltR* et *dltS* codant pour un TCS, en amont de l'opéron *dltABCD*, qui code pour les enzymes impliquées dans le processus de D-alanylation du LTA (voir paragraphe III. A. 1. b.). La protéine DltR fait partie des régulateurs de la famille OmpR. Chez *S. agalactiae*, l'opéron *dlt* est transcrit à partir de deux promoteurs *PdlrR* et *PdlrA* localisés respectivement en amont des gènes

*dltR* et à la fin du gène *dltS* (voir Figure 12)[141]. La transcription se fait principalement à partir du promoteur *PdltR*, qui conduit à la synthèse d'un transcrit de 6.5 kb contenant l'ensemble de l'opéron *dltRSABCD*. En l'absence de D-alanylation (par exemple dans un mutant *dltA*<sup>-</sup>), l'expression des deux promoteurs est induite. L'induction du promoteur *PdltR* dépend de l'action de DltR contrairement à celle du promoteur *PdltA*. Cependant dans le double mutant *dltRA*<sup>-</sup>, l'activité du promoteur *PdltR* n'est pas nulle [141]. Le système DltRS serait impliqué dans le contrôle et dans le maintien du taux de D-alanylation du LTA en fonction des conditions environnementales. Au cours de ma thèse je me suis intéressée à la réponse de *S. agalactiae* à l'absence de D-alanylation, au rôle de DltR dans cette réponse et à la régulation de l'opéron *dlt*.

### 3. CiaRH

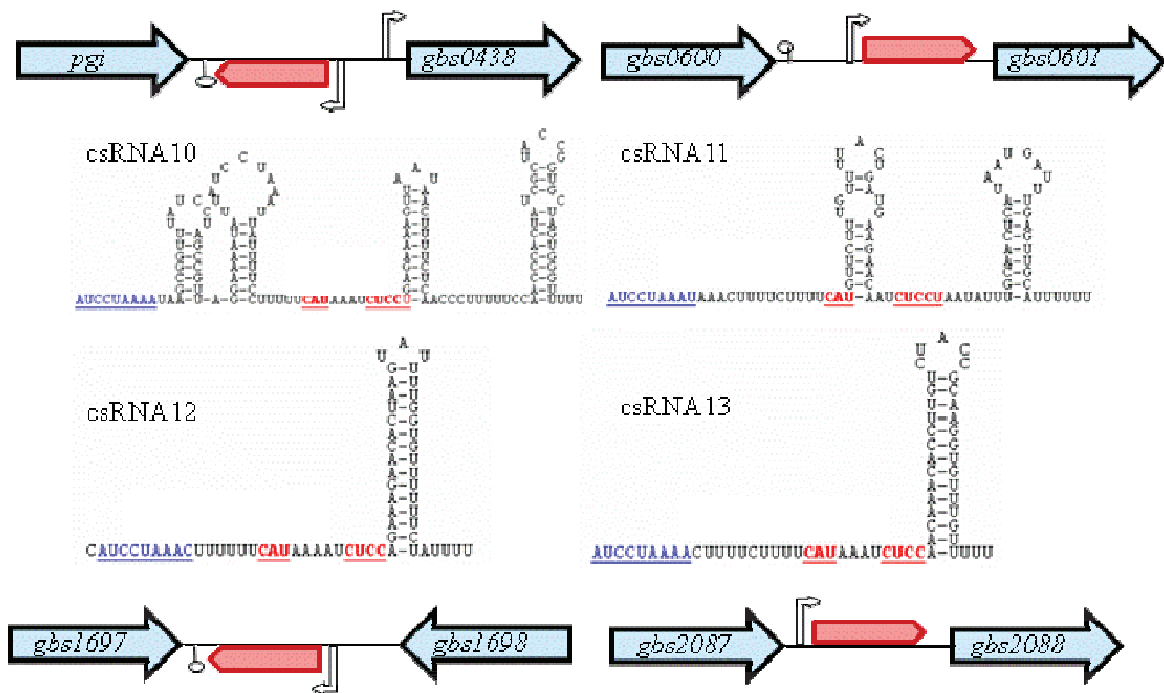
Ce système a initialement été identifié en 1994, chez *S. pneumoniae*, au cours de la recherche de mutants spontanément résistants aux beta-lactamases [207]. La mutation du gène codant pour l'histidine kinase *ciaH* (*ciaH306*; T230P) conduit à une augmentation de la résistance aux beta-lactamases et à une perte de la compétence; en conséquence ce système a été appelé CiaRH (Cia pour « Competence Induction and Altered cefotaxime susceptibility » [207]). Une approche par immunoprécipitation couplée à une hybridation sur puce à ADN a permis de définir 26 sites de fixation de CiaR, localisés dans 18 régions chromosomiques [208]. Ces observations, associées à trois analyses transcriptomiques distinctes, ont conduit à la prédiction d'un "core" régulon de huit loci: *manLMN* (un système de transport du mannose), *axeI* (xylan esterase), l'opéron *cyl* (un système codant pour une toxine cytolytique), l'opéron *malMP* (un système de transport du maltose), l'opéron *dlt*, *htrA* (sérine protéase), l'opéron *lic* (impliqué dans la synthèse de l'acide interchoïque) et *ciaRH* [208-210]. L'analyse des séquences promotrices de ces loci a permis l'identification d'une séquence consensus, TTTAAG-N5-TTTAAG, dix nucléotides en amont de la boîte -10 [211]. Chez *S. pneumoniae*, CiaRH régule l'expression de 5 csRNA (cia dependent small RNA), ces 5 csRNA correspondent à 4 sites de fixation de CiaR observés dans l'expérience d'immunoprécipitation [208]. Par la suite le système CiaRH a également été étudié chez différentes espèces du genre streptocoque: *S. mutans* [212-214], *S. gordonii* [213], *S. suis* [215] et *S. agalactiae* [216].



L'étude réalisée dans la souche COH1 de *S. agalactiae*, montre que le régulateur CiaR est important pour la virulence de *S. agalactiae*: le mutant  $\Delta$ *ciaR* est fortement affecté dans sa survie intracellulaire, dans les neutrophiles et les macrophages et présente une sensibilité accrue aux peptides antimicrobiens, au lysozyme et aux espèces réactives oxygénées. En outre, des expériences de compétition chez la souris montrent que le mutant possède une survie altérée dans le sang. Une analyse transcriptomique indique que l'expression de 6 gènes est diminuée et celle de 1 gène est augmentée. Parmi ces sept gènes figurent *ciaR* (*gbs1020*), une peptidase (*gbs0030*), une protéine de la famille PAP2 (*gbs1195*), une oxydoréductase (*gbs0196*) et une protéine conservée de fonction inconnue (*gbs2003*)[216].

Récemment, une étude bioinformatique basée sur la conservation des protéines CiaR entre les streptocoques (pourcentage d'identité de 80 à 99% par rapport à la séquence de *S. pneumoniae* R6), a permis la prédiction de 40 petits ARNs potentiellement régulés par CiaR, appelés csRNA comme chez *S. pneumoniae*, dans différentes espèces de streptocoques [217]. Tous les génomes de streptocoques étudiés comportent des csRNA, et leur nombre varie de 2 à 6. Chez *S. agalactiae* 4 csRNA ont été prédits (Figure 13). Bien que la fonction de ces csRNA reste inconnue, l'analyse de leurs séquences suggère qu'ils empêcheraient la traduction d'ARNs messagers cibles, en se fixant au niveau de la séquence de reconnaissance du ribosome (la séquence Shine Dalagarno).

Au cours de ma thèse, je me suis intéressée au rôle de CiaRH dans la réponse mise en place par la bactérie en l'absence de D-alanylation, et également de son rôle dans la régulation de l'opéron *dlt*.



**Figure 13 : Structure secondaire et localisation de csRNA chez *S. agalactiae*.**

Réalisée d'après [217]. La structure secondaire a été prédite par le programme Mfold. Les séquences potentiellement complémentaires à la séquence Shine Dalagarno et le site d'initiation de la traduction sont soulignés et indiqués en rouge. Une séquence retrouvée et conservée dans tous les csRNA est soulignée et indiquée en bleu. Les csRNA sont indiqués par des flèches rouges permettant de visualiser la localisation et l'orientation de ces petits ARNs par rapport aux gènes encadrants.

#### 4. RgfBDAC ("Regulatory of Fibrinogen binding")

Ce système a été identifié suite à la recherche de mutants affectés dans leur capacité à se lier au fibrinogène [218]. Les quatre gènes *rgfBDAC* sont transcrits en un seul ARNm. *rgfD* code pour un peptide hormone, *rgfC* une histidine kinase, et *rgfA* un régulateur transcriptionnel. Le phénotype du mutant s'explique par l'augmentation de l'expression du gène *scpB* (C5a peptidase) dans le mutant  $\Delta rgfC$ . Une étude récente montre que RgfAC est également impliqué dans l'activation de la transcription du gène *fbsB* et dans la répression du gène *fbsA* [219]. Le régulateur RovS (voir plus bas) réprime également directement l'expression de *fbsA*. Le niveau de transcription du gène *rovS* n'étant pas modifié dans le mutant  $\Delta rgfAC$ , il semblerait que le contrôle de RgfAC sur l'expression du gène *fbsA* soit indépendant de RovS [219].

Parmi les souches séquencées, l'opéron *rgf* est intact uniquement dans les souches COH1 (ST17), et CJB111 (ST1). Dans les souches H36B (ST6) et A909 (ST7), le gène *rgfA* contient un codon stop en phase et est très altéré dans les autres souches, dont les ST23. Le gène codant pour l'HK manque également dans ces autres souches. Ce polymorphisme dans le locus *rgfABCD* explique en grande partie pourquoi les protéines impliquées dans la liaison au fibrinogène sont essentiellement FbsB dans les souches du ST17 et FbsA dans les souches du ST23. L'implication majoritaire de FbsB dans les souches du CC17, peut s'expliquer par l'inhibition de l'expression du gène *fbxA* par deux mécanismes: tout d'abord par l'effet inhibiteur du système RgfAC, et par l'absence du gène codant pour RogB, un activateur de *fbxA* (voir plus bas) [164, 219].

## B. Les Régulateurs simple (*stand alone*)

### 1. RovS et les régulateurs de la famille Rgg

Les protéines codées par les gènes *gbs0230*, *gbs1555* et *gbs2117* présentent des similarités significatives avec des protéines régulatrices de la famille Rgg, qui contrôlent chez *S. pyogenes* l'expression de nombreux gènes de virulence connus ou putatifs.

Chez *S. agalactiae* la délétion de *rovS* (*gbs1555*) conduit à une augmentation de la liaison de la bactérie au fibrinogène et aux cellules épithéliales pulmonaires A549, suggérant que RovS régule la liaison de la bactérie à la MEC et aux cellules épithéliales de l'hôte [220]. Il a été montré par RT-qPCR et retard sur gel, que RovS régule directement l'expression des gènes *fbxA*, *gbs0230*, *sodA* et *cyl*. Le site de fixation de RovS est une séquence palindromique espacée de 6/7 nucléotides dont le consensus est: AWAAWVHTDAW-N<sub>6/7</sub>-WTKWWAMDWAK. Des sites potentiels de fixation de RovS ont été prédits dans la souche NEM316 par le programme Predict Regulon [221] au niveau des séquences promotrices de 5 gènes du métabolisme central (*gbs0091*, *gbs0871*, *gbs1088*, *gbs1376*, et *gbs2122*), huit gènes de fonction inconnue (*gbs0399*, *gbs0662*, *gbs0729*, *gbs0828*, *gbs0967*, *gbs0980*, *gbs1202*, et *gbs2112*), cinq régulateurs transcriptionnels ou protéines de l'appareil de transcription (*gbs0094*, *gbs0425*, *gbs1381*, *gbs1509*, et *gbs1712*), quatre transporteurs putatifs (*gbs0119*, *gbs1332*, *gbs2007*, et *gbs2017*), et trois gènes codant pour des protéines de surface (*gbs0451*, *gbs1478*, et *gbs1998*). L'ensemble de ces résultats indique que RovS pourrait contrôler l'expression d'un grand nombre de gènes différents chez *S. agalactiae* [220]. Néanmoins la validation expérimentale de ces prédictions n'a pas été réalisée.

## 2. RogB

Le génome de la souche NEM316 contient trois gènes codant pour des régulateurs transcriptionnels de la famille des RofA-Like Protein (RALP) *gbs1479* (*rogB*), *gbs1530* (*rga*) et *gbs1426*. Le gène *rogB* est associé au locus pilus PI-2a et *rga* au locus *secA2*. Le gène *gbs1146* est situé dans le "core" génome et n'est pas associé à des gènes codants pour des composés de l'enveloppe. Chez *S. pyogenes* les régulateurs RofA et Nra contrôlent l'expression de gènes qui codent pour des protéines médiant l'interaction de la bactérie avec la MEC. Chez *S. agalactiae* le mutant de la souche 6313 déficient pour *rogB* est affecté dans sa capacité à se lier au fibrinogène et à la fibronectine. Ce mutant présente également une adhérence réduite aux cellules épithéliales [85], mais n'est pas altéré dans sa capacité d'invasion des cellules eucaryotes [164]. Des analyses par qRT-PCR ont montré que RogB stimule l'expression des gènes *fbsA*, *rogB*, *gbs1478-gbs1477* et réprime l'expression du gène *cpsA*. Le régulateur transcriptionnel RogB semble induire sa propre expression [164]. Le rôle de RogB dans l'expression du locus pilus I-2a a été confirmé dans une étude réalisée dans la souche NEM316 [85]. Dans les souche 2603V/R et JM9130013, le gène *rogB* est inactivé par une délétion d'une base conduisant à un décalage du cadre de lecture, cette inactivation du gène *rogB* explique que la souche JM9130013 n'exprime à sa surface que le pilus I-1 [85, 222].

L'ensemble de ces résultats montre que chez *S. agalactiae* l'expression d'un même facteur de virulence peut être contrôlée par différents systèmes de régulations. Ces systèmes de régulations sont particulièrement abondants chez *S. agalactiae*. Le polymorphisme au niveau de la présence et/ou l'expression de ces systèmes de régulations est un élément important pour expliquer la diversité des souches et éventuellement la différence de virulence entre les souches.

## C. **Adaptation à différents environnements: études transcriptomiques**

Plusieurs études transcriptomiques, publiées récemment, traitent de la réponse mise en place par *S. agalactiae* dans différents environnements, ou suite à des modifications environnementales. Les différents facteurs testés sont regroupés dans le tableau 3, et correspondent entre autre à l'incubation en milieu sanguin, la variation de la température, la croissance en liquide amniotique. L'incubation de *S. agalactiae* pendant 30 ou 90 minutes dans le sang humain conduit à la modification de l'expression de 792 gènes, qui sont pour la majorité (83%) réprimés, tels que l'opéron codant pour la capsule et pour l'antigène B. Cette réponse semble faire intervenir un grand nombre de régulateurs transcriptionnels (28 régulateurs sont différentiellement exprimés). De façon intéressante, parmi les gènes dont l'expression est induite par le sang sont retrouvés les gènes *bibA* (la protéine BibA qui est capable de se fixer à la protéine C4), *fbxA*, *gbs1420* (bsp), *gbs0608* (enolase) et *gapC* (GADPH). Ces transcriptomes montrent que l'exposition au sang induit l'expression de facteurs qui contribuent à la survie dans le sang et favorisent et la dissémination de *S. agalactiae* vers d'autres sites anatomiques [170].

Les variations de température affecteraient essentiellement la transcription des gènes de *S. agalactiae* pendant la phase stationnaire, avec 25% des gènes différentiellement exprimés. L'expression de certains facteurs de virulence, ou protéines de surface, est induite à 40°C. Ce qui est le cas par exemple de l'opéron *cyl*, codant pour l'hémolysine, et de 7 gènes codant pour des protéines à motif LPXTG, dont BibA, ScpB, Srr-1, et des protéines de fonctions inconnues comme Gbs1929, Gbs0428, Gbs1288 et Gbs1403. Au contraire, l'expression d'autres protéines LPXTG est induite à 30°C, ce qui est le cas de *gbs0451* (ScpA) et de *gbs2008*. L'expression des gènes codant pour des protéines impliquées dans le métabolisme des purines et l'homéostasie des ions est également significativement induite à 40°C [223].

Le liquide amniotique est un milieu relativement pauvre en nutriment et est constitué principalement d'eau, d'urée, d'une faible quantité d'acides aminés, de kératine et de protéines [224-226]. La comparaison des réponses transcriptomiques lors d'une croissance en liquide amniotique humain et en TH (Todd Hewith milieu riche), montre que la majorité de différences observées concernent des gènes impliqués dans le métabolisme des acides aminés, des carbohydrates et des nucléotides. De manière surprenante *S. agalactiae* est capable de se diviser

rapidement dans le liquide amniotique et aucune réponse classique au stress n'est observée. Il semblerait que *S. agalactiae* s'adapte facilement au liquide amniotique et ne présente pas de réponse classique au stress. L'expression de plusieurs TCSs est réprimée en condition de croissance dans le liquide amniotique, comme les systèmes Gbs1397/1398, CovRS, Gbs1934, et RgfA. Gbs1397/1398 possède 78% de similarité avec le système SptRS de GAS (Spy874/875), et est requis pour la survie dans la salive, l'expression de facteurs de virulence et l'utilisation de carbohydrates chez GAS [227]. La croissance en liquide amniotique modifie aussi l'expression de facteurs de virulence. En particulier, elle réprime l'expression de l'opéron *cyl*, du gène *lmb*, du locus pilus PI-2a, de l'alpha like protéine *alp2*, et induit l'expression de la capsule et du facteur CAMP [228].

Dans ces trois analyses, le nombre de gène différentiellement exprimé est très élevé (jusqu'à 1078 gènes) rendant l'interprétation des résultats difficile.

**Tableau 3: Liste des analyses transcriptomiques réalisées dans *S. agalactiae*.**

Condition	Souche	Reference
Réponse suite à une incubation dans le sang de 30 min ou 90min	NEM316	[170]
Croissance en TH à 30 et 40°C à différentes étapes de la phase de croissance.	NEM316	[223]
Comparaison croissance en TH et dans le liquide amniotique	NEM316	[228]
Régulation génique au cours de la phase de croissance	NEM316	[229]
Analyse du régulon CovRS	NEM316	[183]
Comparaison du régulon CovRS dans les souches NEM316, 515 et A909	NEM316, 515, A909	[230]
Régulation de facteurs de virulence par CovRS en réponse au pH environnemental	2603V/R	[231]
Analyse du régulon MtaR	COH1	[232]

## V. Éléments génétiques mobiles (EGMs) et évolution du génome chez *S. agalactiae*

### A. Structure composite du génome de *S. agalactiae*

En 2002, les séquences génomiques de deux souches (NEM316 et 2603V/R) ont été publiées [125, 233]. Les analyses des génomes de ces deux souches ont mises en évidence l'organisation composite du génome de *S. agalactiae*. L'étude d'Hervé Tettelin était basée sur l'analyse par CGH de 19 souches de *S. agalactiae* [234]. Alors que celle réalisée au laboratoire se fondait sur la comparaison avec les autres génomes de Streptocoques, publiés à l'époque, et en particulier avec *S. pyogenes* [125]. Cette notion a été étendue par la suite avec l'analyse comparative de huit génomes de *S. agalactiae* [234] et de 75 souches, par la technique de CGH [39]. Les gènes spécifiques de *S. agalactiae* appartiennent à deux groupes. Le premier groupe de gènes spécifiques à *S. agalactiae* représente 468 gènes (pour NEM316 par rapport à *S. pyogenes*) répartis en petites régions sur le core génome. Ces petites régions ne présentent pas de caractéristiques similaires à celles des EGMs et sont potentiellement impliquées dans l'interaction avec l'environnement car elles codent en particulier pour des protéines de surface, ou des systèmes d'utilisation de sucres. Certains de ces gènes variables sont regroupés dans sept régions localisées aux mêmes positions relatives sur les chromosomes, indiquant un remplacement génique par recombinaison au cours de l'évolution [39, 233]. Ces sept régions correspondent: (1) aux gènes codant pour les protéines nécessaires à la synthèse du polysaccharide capsulaire, (2) aux deux régions codant pour des structures de type pilus, (3) au locus *secA2*, (4) au gène codant pour la protéine R5 (également appelée BPS) spécifique de la souche 2603V/R, qui est remplacé par deux gènes codant pour un ABC transporteur (*gbs1400-01*) dans les sept autres souches séquencées, (5) au locus *gbs1740-48*, qui est remplacé par 7 gènes de fonctions inconnues dans les souches A909 et 2603V/R, (6) à la région SAK\_0184-91 codant pour l'antigène  $\beta$  (*bca*), (7) un système à deux composants présent uniquement chez A909 et H36B. Le second groupe contient les gènes regroupés dans des régions chromosomiques correspondant à des éléments génétiques mobiles (EGMs) probables (soit 13 et 14 EGMs, respectivement dans les génomes des souches 2603V/R et NEM316). Au sein de ces EGMs des gènes codant pour des fonctions potentiellement impliquées dans le transfert et la réplication ont été identifiés. De plus, 6 de ces EGMs sont situés à proximité de gènes codant pour des ARNt et

certains présentent un pourcentage en G+C différent du reste du génome. L'ensemble de ces observations suggère que l'acquisition de ces EGMs s'est faite par transfert horizontal.

## B. Les ICEs des EGMs particuliers

Les EGMs sont des segments d'ADN codant pour les protéines nécessaires à leurs propres mouvements intracellulaires et intercellulaires. Les EGMs étaient traditionnellement classés en différentes familles: les bactériophages, qui à l'état lysogène sont intégrés dans le chromosome, les plasmides, qui sont des structures capables de répllication autonome, les intégrons, les transposons, les transposons conjugatifs, récemment renommés ICEs (pour "Integrative and Conjugative Elements"), et les séquences d'insertion (IS). Il a été démontré que les EGMs jouent un rôle important dans le transfert génétique horizontal. Ils participent au transfert d'un grand nombre de gènes codant pour des protéines impliquées dans diverses fonctions cellulaires accessoires, tels que la résistance aux antibiotiques, le catabolisme des xénobiotiques, la pathogénicité et la symbiose. Le maintien de ces EGMs au sein de la bactérie hôte suggère qu'ils lui confèrent un avantage sélectif. Par exemple, la conservation du système ABC osmoprotectant, ainsi que les gènes de la voie de déamination de l'arginine contenus dans les îlots I et XIV de la souche NEM316, pourraient être un avantage dans la résistance aux chocs osmotiques, ainsi que pour la tolérance à l'acidité. D'autres EGMs de *S. agalactiae* sont porteurs de gènes codant pour des protéines impliquées dans les mécanismes de résistances aux métaux lourds ou aux antibiotiques. Comme par exemple, l'ICE\_*Sa2603\_rplL* et l'ICE\_*Sa2603\_tRNA<sup>Lys</sup>* de la souche 2603 V/R, qui codent tous les deux pour un système de résistance aux métaux lourds [235]. L'ICE\_*Sa2603\_rplL* porte également des gènes codant pour des pompes d'efflux aux macrolides Mef(E) et Msr(D) [235].

La résistance à la tétracycline chez *S. agalactiae* est expliquée dans 88% des cas par la présence d'un ICE apparenté à Tn916 au sein du génome de la souche [148]. Le transfert de Tn916 au sein de *S. agalactiae* a été testé par conjugaison [235]. Aucune preuve directe de transfert n'a été obtenue dans les conditions du laboratoire. Il est possible que les Tn916 observés dans les différentes souches aient été acquis: de manière indépendante, ou que le transfert de Tn916 dépende d'une mobilisation par un autre élément conjugatif, ou du transfert de grands fragments chromosomiques par conjugaison [236].



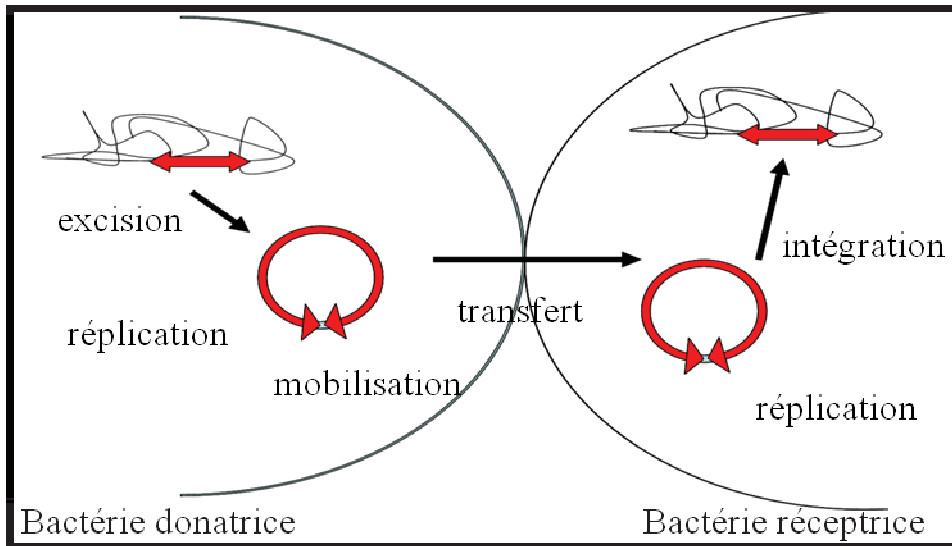
Les ICEs correspondent à des EGMs chimères, ils possèdent une machinerie d'insertion/excision, similaire à celle des bactériophages ou des transposons, et sont également capables de se transférer dans une autre bactérie par conjugaison, comme certains plasmides. Au cours de ma thèse je me suis intéressée à l'étude d'une nouvelle famille d'ICEs, les *TnGBS*.

## **C. Mobilité des ICEs par conjugaison**

Les ICEs correspondent à la combinaison de modules fonctionnels [237]. Les modules des ICEs peuvent être regroupés en différentes catégories fonctionnelles: l'excision/intégration, la mobilisation, le transfert, la régulation, la réplication et les fonctions accessoires [237, 238]. A l'exception des fonctions accessoires, ces fonctions sont toutes en relation avec le processus de conjugaison.

### **1. Description du processus de conjugaison**

Le processus de conjugaison a été étudié chez différentes bactéries gram positif, en prenant comme modèles différents plasmides conjugatifs tels que pCF10, pIP501 et différents ICEs comme Tn916. Le transfert d'un élément génétique par conjugaison nécessite un contact physique stable entre une bactérie donatrice, porteuse de l'ICE, et une bactérie réceptrice. Ce contact permet la mise en place d'un pore assurant le transfert d'ADN à travers les membranes des deux bactéries. Au cours du processus de conjugaison, l'ICE va se trouver sous différents états dans les bactéries donatrice et réceptrice (Figure 14). L'ICE, présent sous forme intégrée dans le génome de la bactérie donatrice, s'excise et se circularise. La forme circulaire est transférée par conjugaison, puis elle s'intègre dans le génome de la bactérie réceptrice.



**Figure 14: Différentes étapes du processus de conjugaison des ICEs.**

**a) Excision de l'ICE, première étape de la conjugaison**

La production d'une forme circulaire de l'ICE est catalysée par une enzyme, codée par un gène porté par l'élément. L'excision des ICEs peut dépendre de l'action d'une tyrosine recombinase (Tn1545, Tn916, pSAM2, ICEBs1), ou d'une sérine recombinase (Tn5397) [239]. Ces différentes enzymes catalysent l'échange de brins, conduisant à la formation d'une forme circulaire d'un élément et à l'intégration de celle-ci [240, 241]. Plusieurs tyrosine et sérine recombinases nécessitent la présence de facteurs accessoires pour contrôler et diriger la réaction de recombinaison (insertion versus excision) [240]. L'excision de Tn916 dans la bactérie donatrice nécessite à la fois l'action de l'intégrase et d'une excisionase. Pour Tn916 et les ICEs proches, la circularisation de l'ICE est requise pour l'expression du module de transfert. En effet, il a été montré que le promoteur contrôlant l'expression du module de conjugaison se situe de part et d'autre du site attI présent dans la forme circulaire de l'élément [242]. L'excision des ICEs apparaît donc comme la première étape nécessaire à leur transfert par conjugaison

**b) Etablissement du contact entre les bactéries**

Chez les bactéries à Gram négatif, le contact est établi par une structure protéique filamenteuse extra-cellulaire appelée *pilus* sexuel, dont la base est ancrée dans la membrane interne par un système similaire aux systèmes de sécrétion de type IV (T4SS Type 4 sécrétion

system). Le pilus croît grâce à la polymérisation de sous-unités de piline, et contacte la cellule réceptrice. Plusieurs pili conjugatifs reconnaissent et se fixent spécifiquement aux lipopolysaccharides des parois bactériennes [243]. Puis le pilus se rétracte grâce à la dépolymérisation des pilines, diminuant ainsi la distance intercellulaire. La rétraction du pilus est supposée être une réponse à la fixation d'une cellule réceptrice à l'extrémité du pilus [244]. Une fois que les cellules donatrices et réceptrices se sont rapprochées, un pore de conjugaison est mis en place. Des études récentes démontrent que la lumière du pilus permet le transport d'ADN simple-brin et son entrée dans la cellule réceptrice [245, 246].

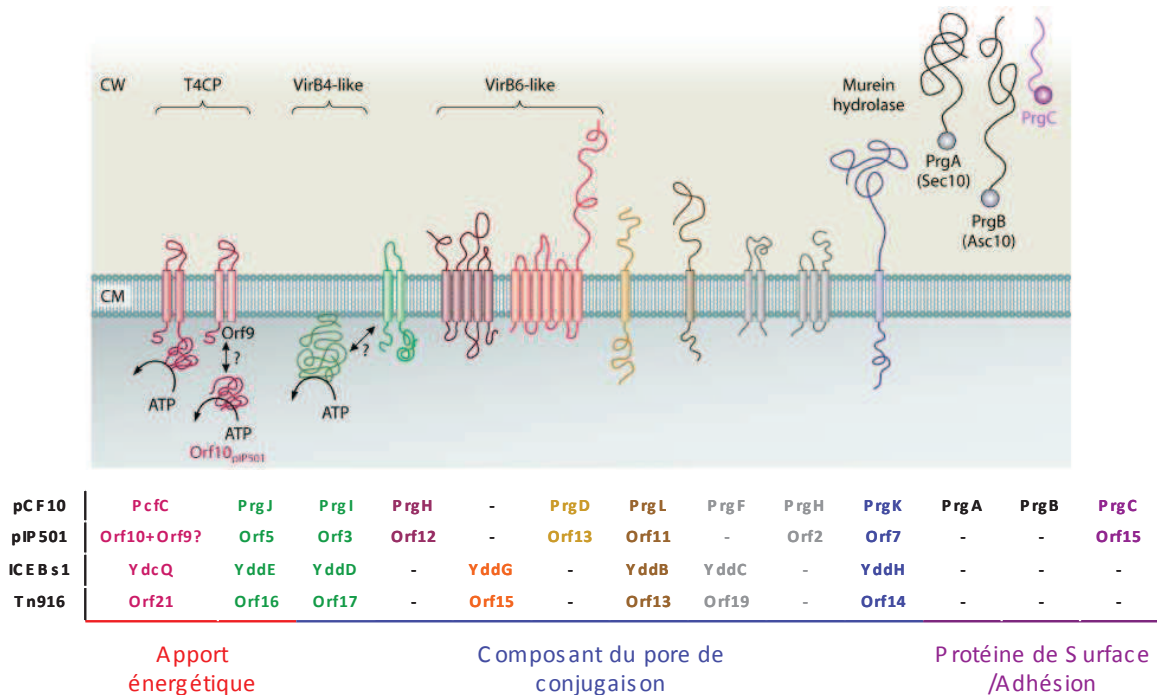
Les plasmides conjugatifs et les ICEs de *Firmicutes* ne codent pas de pili directement impliqués dans le transfert conjugatif. Le contact donatrice-réceptrice serait alors aléatoire. Cependant, certains éléments conjugatifs codent des facteurs d'agrégation pouvant améliorer et favoriser les contacts cellule-cellule [247, 248]. Certains systèmes comme pCF10 et pIP501 codent pour des protéines de surface possédant des motifs LPXTG d'ancrage. Le plasmide pCF10 code pour 3 protéines de surface: PrgA, PrgB et PrgC. PrgA est impliquée dans les fonctions d'exclusion. PrgB contient des motifs internes RGD, impliqués dans l'agrégation entre bactéries, par liaison à l'acide lipotéichoïque, et également dans l'internalisation et la survie dans les cellules épithéliales [249]. La protéine PrgB module l'efficacité de transfert, cependant, elle est facultative pour le transfert d'ADN. Cette protéine ne fait donc pas partie des sous-unités structurelles du système de sécrétion [250]. PrgC et la protéine Orf15<sub>pIP501</sub> possèdent un motif hautement répété Pro-X-Glu/Asp. De tels motifs, riches en proline et en acides aminés chargés négativement, sont fréquemment trouvés chez les bactéries Gram positif dans des protéines de surface, où elles peuvent être impliquées dans l'adhérence de la bactérie aux cellules épithéliales [251, 252], dans l'agrégation des bactéries entre elles [253, 254], dans la formation de biofilms [165, 255]. C'est le cas par exemple de la protéine Sfb (pour fibronectin-binding surface protein) de *S. pyogenes* [256]. Ceci suggérerait que les protéines de type PrgC pourraient conférer un avantage sélectif aux bactéries qui auraient acquis le gène correspondant au sein d'un ICE ou d'un plasmide. Les résultats obtenus au sein de ces différentes études suggèrent que ces protéines, au sein d'éléments conjugatifs, permettraient de conférer un avantage sélectif aux bactéries portant ces ICEs.

### c) Mécanismes de transfert d'ADN

Le mécanisme de transfert d'ADN est un processus commun aux plasmides et aux ICEs. Le mécanisme de transfert des plasmides conjugatifs ayant été plus étudié, ces plasmides servent de modèles pour l'étude du mécanisme de transfert des ICEs [250]. Le module de conjugaison est constitué de deux sous-modules. Le premier sous-module, impliqué dans la mobilisation, permet l'initiation du transfert de l'ADN au cours de la conjugaison. Le second sous-module est impliqué dans la formation et le maintien du pore de conjugaison. Une protéine dite de couplage permet de faire le lien entre ces deux systèmes et confère l'énergie nécessaire au processus de transfert de l'ADN à travers le pore.

Le sous-module de mobilisation permet l'initiation du transfert qui se déroule en plusieurs étapes: la reconnaissance de l'origine de transfert, *oriT*, son clivage [257], la séparation des brins d'ADN et le déplacement d'un des brins vers le pore de sécrétion [258]. On appelle relaxosome le complexe composé de l'origine de transfert, et des protéines nécessaires à ces différentes étapes, c'est-à-dire une relaxase et une ou plusieurs protéines auxiliaires. Généralement, l'origine de transfert correspond à une séquence de 40 à 500 pb présentant des répétitions directes et indirectes, qui permettent la fixation des protéines impliquées dans le transfert d'ADN [248]. Ce site d'initiation de transfert est clivé par la relaxase. Les relaxases sont structurellement apparentées aux protéines initiateuses de la réplication par cercle roulant, et catalysent des réactions biochimiques similaires. Par une analyse phylogénétique, les relaxases ont été classées en sept familles: MOB<sub>F</sub> (plasmide F), MOB<sub>H</sub> (SXT), MOB<sub>Q</sub> (pIP501), MOB<sub>C</sub> (pAD1), MOB<sub>P</sub> (pCF10), MOB<sub>V</sub> et les relaxases apparentées à Tn916 (Tn916, *ICEBs1*, *ICESt1*) [259]. Dans les différentes familles de relaxase, l'acide aminé catalytique responsable de l'attaque nucléophile permettant la coupure simple-brin de l'ADN, au niveau de l'origine de transfert, est un résidu tyrosine [260, 261]. Cette réaction conduit à la formation d'une liaison covalente entre le résidu tyrosine de la relaxase et l'extrémité 5' de l'ADN simple-brin. Bien qu'un gène codant une relaxase putative soit présent chez la plupart des ICEs, les relaxases et les coupures simple-brin qu'elles catalysent au niveau de l'*oriT* n'ont été caractérisées que pour un petit nombre d'éléments dont Tn916, Tn5252 et *ICEBs1* [257, 262, 263]. Alors que la relaxase NicK de *ICEBs1* est capable de catalyser le clivage au niveau de l'origine de transfert en absence d'autres

facteurs [262], d'autres relaxases nécessitent des facteurs accessoires permettant leur recrutement au niveau de l'origine de transfert, ce qui est le cas de la relaxase PcfG du plasmide pCF10 [264]. Il a été montré pour Tn916, que la présence de l'intégrase modifie l'activité de la relaxase (Orf20). En présence de l'intégrase, la relaxase coupe efficacement le brin d'ADN au niveau de l'origine de transfert. En l'absence de l'intégrase, l'Orf20 agit comme une endonucléase et coupe les deux brins d'ADN au niveau de différents sites de l'*oriT* enrichis en GC [263]. Le monomère de relaxase lié à l'ADN, serait reconnu par la protéine de couplage permettant ainsi son transfert à travers le pore de conjugaison. Un second monomère de relaxase permettrait quant à lui, grâce à son activité hélicase, la réplication de l'élément par cercle roulant chez la cellule donatrice [265, 266]. Le transfert simple-brin n'a été réellement démontré que pour un seul ICE, Tn916 [267].



**Figure 15 : Composants de l'appareil de transfert des plasmides pCF10 et pIP501 et des deux ICEs ICEBs1 et Tn916.**

D'après [250]

Les protéines de couplage servent de lien entre les deux machineries impliquées dans le transfert et fournissent l'énergie nécessaire au transfert de la molécule d'ADN, grâce à leur activité ATPasique [265]. Elles appartiennent à la famille des protéines SpoIIIE/FtsK, qui sont des transporteurs d'ADN impliqués dans de nombreux processus de translocation d'ADN entre compartiments cellulaires [268], et sont apparentées aux protéines VirD4 des Systèmes de

Sécrétion de type IV des bactéries Gram négatif. Dans le cas de pCF10, il a été montré que PcfC interagit avec la relaxase PcfG, la protéine accessoire PcfF et le plasmide sous forme processée [269]. Au niveau de l'appareil de transfert, PcfC interagit avec la protéine PrgH (VirB6-like) et la protéine PrgJ (VirB4-like) [250]. Même si cela n'a pas été montré pour pCF10, certaines protéines de couplage interagissent également avec des protéines codant pour des hydrolases du peptidoglycane, ce qui est le cas pour le plasmide pIP501 de *S. agalactiae* [270]. L'implication dans le transfert conjugatif des protéines de couplage putatives, identifiées chez les différents ICEs, n'a jamais été étudiée, bien qu'un gène codant une protéine de cette famille soit toujours présent.

Le sous-module de transfert code pour l'ensemble des protéines impliquées dans la formation et le maintien du pore de conjugaison, ainsi que pour des protéines permettant la dégradation du peptidoglycane. Certaines de ces protéines sont homologues aux protéines VirB1, VirB4 et VirB6 des Systèmes de Sécrétion de type IV (T4SS) des bactéries Gram négatif [238, 248].

Les protéines composant le pore de conjugaison ont pour caractéristique, presque générale, de posséder un ou plusieurs (de deux à sept) domaines transmembranaires. Quand ce domaine est unique, il peut être N-terminal ou central. Une ATPase de type VirB4 (PrgJ, Orf15, YddE, Orf16) est trouvée pour tous les plasmides (pCF10 et pIP501) et les ICEs (ICEBs1 et Tn916) étudiés chez les bactéries à Gram positif. Une deuxième protéine de type VirB4 (PrgI, Orf3, YddD, Orf17), possédant 2 ou 3 domaines transmembranaires, est également systématiquement retrouvée. L'appareil de transfert compte également plusieurs protéines apparentées à VirB6 (telles que PrgH, Orf12, YddG, Orf15) présentant un grand nombre de domaines transmembranaires et permettant la formation d'un tunnel à travers la membrane bactérienne. Dans le cas de Tn916, de ICEBs1 et de pCW3, les protéines VirB6-like possèdent également une grande extension en C-terminal. Ce domaine C-terminal est important pour la fonction de cette protéine. En effet, pour pCW3 la délétion de ce domaine abolit le transfert de ce plasmide [271].

Les protéines impliquées dans la dégradation du peptidoglycane (telles que prgK, Orf7, YddH, Orf14) [272] sont des muréines hydrolases appartenant à différentes classes, selon leurs cibles au niveau du peptidoglycane. Bien que ces hydrolases soient facultatives pour la formation

du pore de conjugaison à travers la paroi des bactéries Gram négatif, elles semblent essentielles pour la formation du pore de conjugaison chez les Gram positif [250].

En 2010, une analyse phylogénétique, de l'ensemble des protéines présentant des domaines VirB4, a révélé que ces protéines appartiennent majoritairement à quatre groupes. Ces différents groupes permettent de définir quatre groupes de modules de transfert nommés MPF<sub>T</sub>, MPF<sub>F</sub>, MPF<sub>I</sub> et MPF<sub>G</sub> [273].

#### d) **Intégration versus réplication**

Après transfert d'une copie complète de l'élément sous forme simple-brin, un second résidu tyrosine de la relaxase catalyserait la re-circularisation de l'élément transféré [261], qui peut alors être maintenu soit par réplication pour les plasmides, soit par intégration dans le chromosome de la bactérie réceptrice pour les ICEs. La majorité des ICEs ne sont pas capables de se maintenir durablement sous forme circulaire, et ne sont transmis à la descendance que sous une forme intégrée [274]. L'intégration de l'ICE dans la bactérie réceptrice fait intervenir les intégrase de type sérine [239] ou tyrosine recombinases [275].

Certains éléments possèdent également un module de réplication, en plus du module d'intégration, permettant la stabilisation de l'élément dans l'hôte. L'*ICEHin1056* peut être trouvé dans certains cas sous une forme uniquement intégrée, et dans d'autres cas il peut être observé à la fois sous forme intégrée, et sous forme libre en multicopie dans la cellule [276]. Lorsqu'une cellule, ne possédant que la forme intégrée de l'*ICEHin1056*, est utilisée comme donatrice, les transconjugants obtenus possèdent non seulement la forme intégrée de l'élément, mais également de nombreuses formes circulaires [276-278]. La réplication de l'ICE sous forme libre semble donc suivre son transfert. La présence de copies libres de l'ICE pourrait favoriser son transfert vers les cellules adjacentes. Un comportement similaire a été décrit pour l'ICE pNOB8 de *Sulfolobus*. Cet élément est généralement présent sous forme de 2 à 5 copies circulaires, coexistant parfois avec une copie intégrée [279, 280]. Toutefois, pendant les quelques générations suivant le transfert, cet élément est présent à près de 30 copies dans la cellule réceptrice [280]. Cet état à fort nombre de copies n'est pas durable, puisqu'après quelques divisions cellulaires, l'élément n'est plus présent qu'en 2 à 5 copies [279]. La réplication implique une primase codée soit par le génome de la bactérie réceptrice, soit par l'élément

conjugatif chez la donatrice puis transférée vers la réceptrice [281-283]. Les primases sont des enzymes qui permettent l'initiation de la synthèse d'ADN en synthétisant une amorce d'ARN complémentaire à l'ADN simple-brin à répliquer.

Il est à noter que pour certains ICEs, une étape de réplication pourrait également précéder le transfert. C'est le cas de l'élément conjugatif pSAM2 de *Streptomyces ambofaciens* qui est répliqué après son excision dans la cellule donatrice [284], et également avant son intégration dans le génome de la bactérie réceptrice. La copie nouvellement transférée se réplique ensuite, avant de se répandre dans le mycélium de la bactérie réceptrice, puis de s'intégrer dans son génome [285]. A la fin du transfert conjugatif, l'ICE ne subsiste que sous une forme intégrée en monocopie. La réplication de l'ICE après son transfert assure ici l'envahissement du mycélium receveur.

Ces différentes données suggèrent donc fortement, que certains éléments peuvent se maintenir sous forme libre par autoréplication, dans certaines conditions. Ces éléments conjugatifs apparaissent donc à la fois comme des éléments intégratifs et des plasmides.

## 2. Le spectre de transfert

La plupart des éléments testés se transfèrent au minimum au sein de l'espèce ou du genre dans lesquels ils ont été isolés. Le spectre de transfert des éléments conjugatifs est variable. Par exemple, certains éléments comme le plasmide pMG1 d'*Enterococcus faecalis* ne se transfèrent qu'au sein de la même espèce ou entre des espèces proches [286]. D'autres plasmides conjugatifs présentent des spectres de transferts très larges. Les plasmides pIP501 et pAM $\beta$ 1, isolés respectivement de *S. agalactiae* et *E. faecalis*, sont capables de se transférer au sein d'un large éventail de firmicutes [287, 288]. De même chez les ICEs, le spectre de transfert est très variable; ainsi le transfert de l'ICE pRS01 de *L. lactis* n'a été obtenu qu'au sein de cette espèce [289]. Au contraire, Tn916 peut se transférer d'*E. faecalis* vers un grand nombre de firmicutes, de diverses protéobactéries et s'y maintenir [290].

## 3. Immunité de conjugaison



Le transfert de certains ICEs ou plasmides peut être affecté par la présence d'un élément identique ou apparenté, suggérant un mécanisme d'immunité. Cette immunité implique des protéines codées par l'élément génétique mobile, limitant l'acquisition par la bactérie hôte d'un élément apparenté proche, ou d'une seconde copie de cet élément. L'immunité peut faire intervenir différents mécanismes tels que: (1) l'inhibition de la formation du pore entre les bactéries, donatrice et réceptrice (exclusion de surface) Un des exemple le plus étudié est celui de la protéine TraT du plasmide F [291, 292]. (2) L'inhibition de l'entrée de l'ADN dans la bactérie réceptrice (exclusion d'entrée) déjà porteuse du plasmide. Cette inhibition est liée à l'action de la protéine TraS dans le cas du plasmide F [293]. La protéine TraS de la bactérie réceptrice interagit avec la protéine TraG dans la membrane de la bactérie donatrice [294, 295]. (3) L'inhibition de l'intégration (immunité de cible) d'abord décrite pour les bactériophages dont le bactériophage Mu [296, 297], elle a également été décrite pour certains transposons codant des transposases à motifs DDE, dont le transposon Tn7 [298-300]. (4) Un dernier type d'immunité identifié au préalable dans les prophages, dont le phage  $\lambda$ , est liée à la présence d'un répresseur cI ou cI-like [301, 302]. Les répresseurs cI et cI-like, respectivement codés par les prophages I et TP901-1, sont responsables d'une immunité de surinfection vis à vis des phages étroitement apparentés. Ces deux types de protéines sont responsables du maintien des prophages sous forme lysogène, en réprimant directement ou non la quasi-totalité des gènes du phage, et en induisant leurs propres transcriptions [301]. Certains ICEs comme l'*ICEBsI* de *B. subtilis* codent également pour un répresseur de type cI-like (ImmR) responsable d'une immunité de surconjugaison [303]. La fréquence de transfert d'*ICEBsI* est réduite d'un facteur  $10^3$  si la bactérie réceptrice contient déjà une copie de cet élément ou du gène *immR*.

#### 4. Régulation du processus de conjugaison

Bien que de plus en plus d'ICEs soient décrits, la régulation de leurs excisions et de leurs transferts n'a été que peu étudiée. Cette régulation est complexe puisqu'elle permet le couplage entre toutes les étapes de la mobilité: l'excision, le transfert, la réplication et l'intégration. Les mécanismes de régulations impliqués semblent néanmoins variés, ainsi que les stimuli induisant la conjugaison. Un même ICE peut répondre à des stimuli de natures différentes ce qui est le cas de l'*ICEBsI* de *B. subtilis*, dont la conjugaison peut être induite par des agents endommageant l'ADN, l'entrée en phase stationnaire ou par un mécanisme de quorum sensing [303-305].

### a) **Induction de la conjugaison par les conditions environnementales** **exemple de Tn916**

Le transfert de plusieurs ICEs est activé dans les conditions environnementales dans lesquelles la possession de l'élément est avantageuse. Ainsi, l'expression des gènes de transfert de Tn916 est induite par la tétracycline, antibiotique contre lequel l'ICE code un gène de résistance [242]. La présence de tétracycline augmente la fréquence de transfert d'un facteur 10 à 100, alors qu'aucunes modifications de la fréquence d'excision ne sont observées [306]. La fréquence de transfert d'un autre ICE codant pour un gène de résistance à la tétracycline, l'ICE CTnDOT est également induite par cet antibiotique [307, 308].

### b) **Régulation de la conjugaison par un système bistable: cas de ICE<sub>clc</sub>**

D'autres éléments sont régulés par un système bistable similaire à celui du phage  $\lambda$ . La régulation du phage  $\lambda$  oscille en permanence entre deux états stables (les cycles lysogène et lytique), ce type de régulation étant qualifié de bistable. Les mécanismes de régulation de plusieurs ICEs constituent des systèmes bistables. Le régulateur InR, codé par l'ICE<sub>clc</sub> de *P. knackmussi*, code pour une protéine activatrice contrôlant l'expression du gène de l'intégrase de l'ICE. ICE<sub>clc</sub> n'est sous forme circulaire en phase exponentielle de croissance que dans 3 à 5% des bactéries [309]. Cette proportion d'éléments actifs est similaire à la fréquence de transfert maximale de l'élément d'environ  $4 \times 10^{-2}$ , suggérant que toutes les cellules activées engagent un transfert conjugatif efficace [309]. L'expression de l'intégrase (*int*) de l'ICE<sub>clc</sub> est également régulée par la circularisation de l'ICE. Dans l'état intégré, l'intégrase est transcrite à partir du promoteur *Pint*, un promoteur faible et sous le contrôle du régulateur InrR. La circularisation de l'ICE conduit au positionnement d'un promoteur constitutif fort, *Pcirc*, en amont du gène *int*, augmentant son expression, permettant la réintégration dans le génome de la bactérie donatrice, ou, après conjugaison son intégration dans la bactérie réceptrice [310, 311].

### c) **Induction de la conjugaison par le système SOS: cas de SXT et ICEBs1**

Le cycle lytique du phage  $\lambda$  est induit par des dommages de l'ADN, au travers d'un répresseur de type CI. Les répresseurs de type CI appartiennent à la famille des autopeptidases.

Ils comptent notamment parmi eux le répresseur LexA du système SOS. De nombreux ICEs codent également pour des répresseurs de type CI. Par exemple, dans le cas de SXT, SetR inhibe l'expression des gènes activateurs de l'excision et du transfert *setC* et *setD*, et active sa propre transcription [312]. Lorsque des rayonnements ultraviolets ou des mutagènes (mitomycine) endommagent l'ADN, des fragments d'ADN simple-brin s'accumulent lors de la réplication. La protéine RecA forme un complexe avec l'ADN simple-brin et se retrouve alors activée (RecA\*). RecA sous forme activée induit la fonction autopeptidase des répresseurs de la famille CI, dont SetR [312]. Ceci entraîne la séparation des domaines HTH N-terminal et C-terminal qui permettaient l'oligomérisation du répresseur et conduit à son inactivation. SetR sous forme clivée ne réprime plus les gènes *setC* et *setD*. L'expression de *setC* et *setD* conduit à l'excision et au transfert de l'ICE, à une fréquence 300 fois supérieure environ à celle observée dans les conditions non activées [313]. SetC et SetD augmentent aussi l'expression d'autres gènes non requis pour le transfert, dont les gènes du système toxine-antitoxine MosA et MosT [314].

Le système SOS induit également le transfert de l'ICE*BsI*. Un traitement à la mitomycine induit l'excision et le transfert de l'ICE*BsI* de façon dépendante de RecA [304]. Cependant, le mécanisme d'activation diffère de celui de SXT. ICE*BsI* code pour un répresseur ImmR, qui inhibe l'expression de l'excisionase (*xis*) et des gènes en aval. Suite à l'induction du système SOS, le clivage du répresseur fait intervenir la métallopeptidase ImmA, codée par le gène en aval de *immR* [305]. Des homologues de ImmA et ImmR ont été identifiés dans différents EGMs de bactéries Gram positif [305]. Le quorum sensing joue aussi un rôle important dans la régulation du transfert de l'ICE*BsI*. ICE*BsI* code pour un régulateur RapI (pour regulator phosphatase I), qui contrôle l'expression de nombreux gènes de l'ICE dont l'excisionase. L'expression du gène RapI est bloquée par un peptide inhibiteur (PhrI) [304]. Dans les conditions où les cellules porteuses de l'ICE sont en faible proportion par rapport à la population totale, le niveau de PhrI est faible, et le transfert est induit.

#### **d) Régulation par la présence de cellules réceptrices, cas particulier des plasmides à phéromones**

Chez *Enterococcus*, il a été montré, depuis 1978 [315], l'existence d'une classe d'appareils de transfert originale, qui répond à la production de phéromones. La plupart des souches d'entérocoques produisent de multiples petits peptides extracellulaires hydrophobes (sexe phéromones), qui peuvent agir comme signal de communication inter-bactérienne. Ces

petits peptides hydrophobes induisent l'expression des gènes de transfert. Chaque système de transfert est activé par différentes phéromones. Afin d'éviter l'auto-activation de la bactérie hôte en absence de bactérie réceptrice, un système régule finement la production de phéromones endogènes. Ces systèmes permettent un transfert vers un groupe limité d'espèces bactériennes [316, 317].

Parmi les plasmides à phéromones, les plus largement étudiés sont pAD1 [318], pPD1, pAM373, et pCF10. Dans le cas de pCF10, la phéromone (cCF10) est codée par le gène *ccfA* sur le chromosome, alors que les éléments impliqués dans la détection du signal et la mise en place de la réponse sont localisés sur le plasmide [317]. La cellule donatrice, en présence de phéromones, produit une structure à sa surface appelée substance d'agrégation (Asc10) [319]. Cette substance, avec PrgA, participe à l'agrégation des bactéries donatrice et réceptrice. Après acquisition du plasmide pCF10 par la bactérie réceptrice, il y a diminution de la production de la phéromone cCF10, et production d'un peptide inhibiteur iCF10 [320]. Il a été montré que l'état d'induction du système, dans la bactérie donatrice, est dépendant du ratio des deux peptides signaux. L'activité de la phéromone endogène est contrôlée par deux éléments. Tout d'abord par le peptide inhibiteur (iCF10) qui empêche l'activation du système, la phéromone et le peptide inhibiteur interagissant tous deux avec la protéine régulatrice (PrgX), qui régule l'expression des gènes de transfert. Le second système de régulation du taux de phéromones endogènes passe par une protéine transmembranaire (PrgY) [321]. Cette protéine reconnaît la phéromone mature à la surface de la bactérie et inactive sa fonction [316].

**Tableau 4: Les différentes étapes du processus de conjugaison de pAD1 et pCF10 ainsi que les protéines impliquées dans ces différentes étapes.**

(d'après [322])

	Protéines et gènes impliqués dans les différentes étapes de la conjugaison	
	pAD1	pCF10
1. Sécrétion de la phéromone chromosomique par la réceptrice potentielle	cAD1 ( <i>cad</i> )	cCF10 ( <i>ccfA</i> )
2. Reconnaissance et internalisation de la phéromone par la bactérie donatrice	TraC	PrgZ
3. Induction du système de conjugaison	TraA	PrgX
4. Production de la substance d'agrégation (AS) au niveau de la cellule donatrice	Asa1 ( <i>asa1</i> )	Asc10 ( <i>prgB</i> )
5. Agrégation entre les bactéries donneuse et réceptrice	Sea1 ( <i>sea1</i> )	Sec10 ( <i>prgA</i> )
6. Initiation du transfert	<i>oriT1</i> , <i>oriT2</i> , TraX, TraW	<i>oriT</i> , PcfG
7. Répression des fonctions de conjugaison après acquisition dans la cellule réceptrice.	iAD1 ( <i>iad1</i> )	iCF10 ( <i>prgQ</i> )

## 2. RESULTATS

### I. Réseaux de régulation associés aux composants de la paroi

La technologie des microarrays permet en une seule étape d'avoir un panorama complet de la concentration relative d'ARN pour chaque gène et d'en inférer le niveau d'expression [323]. La comparaison des profils d'expression entre différentes conditions de croissance d'une bactérie permet donc de mettre en évidence à l'échelle globale l'adaptation à ces différentes conditions. Les analyses transcriptomiques permettent aussi de mieux caractériser la fonction de gènes régulateurs par comparaison entre les profils d'expression des souches mutantes et de la souche sauvage. Dans la première partie de la thèse, j'ai analysé certains des réseaux impliqués dans la régulation de la synthèse de composants de la paroi. Pour cela j'ai utilisé deux stratégies: l'analyse des effets de perturbations touchant la paroi bactérienne (niveau de D-alanylation du LTA, interaction avec des peptides antimicrobiens par exemple), et l'analyse de mutations dans des gènes régulateurs candidats. Ces régulateurs candidats ont été choisis notamment pour leurs proximités physiques au niveau chromosomique avec des gènes codant pour des protéines de surface. Ces analyses ont été réalisées dans le cadre de différentes collaborations avec des équipes de l'Institut Pasteur (unité de Biologie des Bactéries Pathogènes à Gram-Positif dirigée par Patrick Trieu-Cuot), de l'Hôpital Cochin (CNR Streptocoques, INSERM 1016, dirigé par Claire Poyart), et de l'INRA de Jouy en Josas (UMR1319 Micalis avec Philippe Gaudu). La majorité des conditions testées affectent l'interaction entre la bactérie et son environnement, car elles modifient la présentation des structures de la paroi, et de protéines de surface. Ceci m'a permis d'obtenir une base de données de 34 expériences de transcriptome impliquant 11 projets (Tableau 8). Le plus souvent les perturbations étudiées entraînent la modification de l'expression d'un grand nombre de gènes (>100) résultant d'effets directs et indirects. La combinaison de différentes expériences permet, dans une certaine mesure, de décomposer la réponse afin de mieux la décrire et la comprendre. C'est ce que j'ai réalisé, dans l'interprétation de l'effet de l'absence de la D-alanylation, qui est due à la délétion du gène *dltA*.

## A. **Matériels et méthodes**

La partie *matériels et méthodes* qui suit, correspond aux expériences qui ne sont pas présentées dans les publications.

### 1. **Construction des mutants insertionnels des gènes *ItaS* et *gbs1493***

Des fragments d'ADN, correspondant à la partie centrale des gènes *ItaS* et *gbs1493*, ont été obtenus par PCR à partir de l'ADN de la souche NEM316, en utilisant respectivement les amorces Eco\_gbs1451-Bam\_gbs1451, Eco\_gbs1493-Bam\_gbs1493, dont les séquences sont disponibles dans la table 5. Ces fragments ont été clonés entre les site *Bam*HI et *Eco*RI du plasmide pG1, un plasmide dérivé de pG+host5 avec une origine de répllication thermosensible, et codant pour la résistance à l'érythromycine [324]. Après vérification par séquençage, les plasmides ont été introduits par électroporation dans la souche NEM316. La présence du plasmide a été confirmée par PCR à partir des colonies avec les oligonucléotides P et RP. L'intégration du plasmide pG1 par recombinaison homologue a été sélectionnée, en élevant la température de culture à 40°C comme décrit dans l'article n°1. Après étalement sur boîte, les colonies résistantes à l'érythromycine, obtenues à 37°C ont été mises en culture pour extraction de leur ADN génomique, avec le kit DNAeasy de Qiagen, en utilisant le protocole recommandé pour les bactéries à Gram positif. L'insertion du plasmide pG1 a été confirmée par PCR et par séquençage.

Tableau 5: Liste des oligonucléotides utilisés pour la première partie de la thèse

Nom des oligos	Séquence	
Eco_gbs1451	AACTAGAATTCATTATGCCAAACCCAATCCA	mutant synthèse LTA
Bam_gbs1451	GAACAGGATCCCGGTTGTTGCATGTGCTCTA	mutant synthèse LTA
Eco_gbs1493	AACTAGAATTCTAAAGTCGGTGGCTTATTGT	mutant synthèse AgB
Bam_gbs1493	GAACAGGATCCGCCCATGAATATAAGCAAAG	mutant synthèse AgB
P	CGCCAGGGTTTTCCAGTCACGAC	vérification des constructions
RP	AGCGGATAACAATTTACACAGGA	vérification des constructions

### 2. **Analyses transcriptomiques**

Les analyses transcriptomiques ont été réalisées en utilisant des puces dessinées au laboratoire à partir des séquences génomiques des souches NEM316 et COH1, et contenant 3 sondes par gène pour les 2134 gènes de NEM316 (*gbs*) et des sondes intergéniques dans les deux orientations, ainsi que des sondes pour les gènes annotés de la souche COH1 (SAN\_). Les puces ont été produites par Agilent. Pour chaque souche mutante ou sauvage les ARNs ont été extraits, à partir de trois cultures indépendantes, par la technique au trizol [325]. Puis les ARNs ont été reverse transcrits, en utilisant le kit Superscript indirect cDNA (Invitrogen), et marqués avec les cyanines Cy5 et Cy3 (Amersham Biosciences) selon les recommandations du fabricant. Lors de l'hybridation, des échanges de marquage par les cyanines ("dye swap") ont été réalisés afin de limiter les biais liés au marquage. La fluorescence des différentes sondes des puces a été quantifiée à l'aide du scanner Axon 4000B, connecté au programme d'acquisition d'image GenePix Pro 6.0 et son logiciel d'analyse. Un pré-scan a été effectué pour vérifier globalement l'efficacité de l'hybridation, et pour déterminer les paramètres d'acquisition de l'image. Après génération des images, les spots avec des intensités de fluorescence anormales, soit trop proches du bruit de fond, soit saturés, ont été éliminés de l'analyse. Pour les analyses, aucun bruit de fond n'a été retiré. Les données extraites des différents champs ont été normalisées en utilisant une procédure de normalisation globale LOESS [326], disponible dans le package "Bioconductor package marray" sur le programme R (<http://www.r-project.org>), pour corriger les biais liés au marquage. Pour déterminer si les gènes sont différentiellement exprimés, un t-test, en séries appariées, a été réalisé en utilisant la méthode de Benjamini et Yekutieli pour corriger les p-values [327]. La différence d'intensité pour chaque sonde entre les deux conditions est considérée comme significative lorsque la p-value est inférieure à 0.05. Puis pour chaque gène un ratio moyen a été calculé à partir de celui obtenu avec les trois sondes. Les gènes dont le ratio moyen est inférieur ou égal à 0.5 ou supérieur ou égal à 2 ont été considérés comme pertinents et ont été conservés pour l'analyse.

### **3. Classification des gènes (Clustering)**

Au travers des différentes collaborations, 34 analyses transcriptomiques ont été réalisées dans la souche NEM316. Parmi ces 34 analyses, seules les 26 comparaisons réalisées contre la souche NEM316, sauvage, en culture statique en début de phase exponentielle (DO 0,3 - 0,4) ont été conservées pour la classification. Les gènes pour lesquels la valeur du ratio était absente



lors d'une expérience ont été retirés, ce qui a conduit à garder 1907 gènes pour l'analyse. La majorité des gènes non pris en compte correspond à des gènes d'éléments génétiques mobiles qui sont souvent très faiblement exprimés. 1233 gènes étaient différentiellement exprimés d'un facteur 2 dans au moins une des 26 expériences. Ce nombre étant important, j'ai décidée de ne sélectionner pour la classification que les 636 gènes pour lesquels une expression différentielle de 3 a été observée dans au moins un des transcriptomes. Les valeurs des ratios de ces 636 gènes, dans les 26 comparaisons, ont été utilisées pour faire une classification hiérarchique afin de visualiser et de regrouper les gènes régulés de façon similaire selon les expériences. Cette classification hiérarchique ascendante a été réalisée via le programme Cluster 3 [328], grâce à une approche par saut maximal ("complete linkage"), ce qui a tendance à produire des classes spécifiques en regroupant des classes très proches. Le résultat de cette classification a été visualisé en utilisant le programme Java Treeview [329].

#### **4. Recherche de motifs**

La recherche des sites de fixation des régulateurs transcriptionnels chez *S. agalactiae* NEM316 a été réalisée en utilisant la fonction "Search Pattern", disponible sur le site <http://genolist.pasteur.fr/SagaList/index.html>, en prenant pour la recherche soit les 200pb en amont du codon d'initiation des gènes, soit l'ensemble des régions inter-géniques. Les séquences, utilisées pour cette recherche, correspondent aux consensus des séquences cibles des différents régulateurs au sein d'autres espèces de streptocoques. La recherche des sites de fixation, de ces mêmes régulateurs pour des gènes homologues, chez les autres espèces du genre streptocoque a été faite en utilisant le programme RegPredict [330]. La recherche des sites de fixation a été réalisée à partir des séquences génomiques des souches TIGR4, M1, et UA159 respectivement pour *S. pneumoniae*, *S. pyogenes* et *S. mutans*, et correspondant aux numéros d'accèsion NC\_003028.3, NC\_002737.1, NC\_004350.2.

#### **5. Effet de la colistine**

Les concentrations sub-inhibitrices de colistine pour *S. agalactiae* ont été déterminées par des expériences de croissance de la souche NEM316 en présence de différentes concentrations

de colistine, sur une gamme de 0 µg/mL à 250 µg/mL (0, 10, 20, 30, 40, 50, 60, 80, 100, 125, 150, 175, 200 et 250µg/mL). La souche NEM316 a été mise en culture à DO 0,01 à partir d'une préculture sur la nuit. A DO 0,05 la culture a été divisée et la colistine a été ajoutée afin d'obtenir les différentes concentrations finales de la gamme. L'analyse transcriptomique a été réalisée à partir de bactéries récupérées à DO 0,4-0,45 en présence de 60 µg/mL de colistine.

## **6. Analyse des données d'identification des sites d'initiation de la transcription (TSS)**

Le protocole choisi pour l'analyse des données de séquençage repose sur la comparaison des résultats obtenus pour les extrémités 5' des ARNs traités ou non par la Phosphatase acide du tabac ou TAP. La TAP permet la conversion des extrémités 5' tri-phosphates des ARNs bactériens "natifs" (résultant directement de l'étape de transcription et non dégradés) en extrémités 5' mono-phosphates compatibles avec la ligation d'adaptateurs. En absence de traitement par la TAP, le séquençage donne accès à tous les ARNs (ARN ribosomiques, tRNAs, ARNm en cours de dégradation ou ayant subi une maturation) à l'exception de ceux dont l'extrémité 5' est native. Les séquences obtenues après traitement par la TAP et réalisées en parallèle permettent d'avoir accès, en plus, aux ARNs "natifs". La comparaison du nombre de lectures (séquences) obtenues avec les deux protocoles permet de dégager, après analyse statistique, les séquences correspondant aux extrémités natives des transcrits.

## **B. Résultats et discussions**

### **1. Effet sur le profil d'expression de perturbations touchant des structures non protéiques de la paroi (publication n°1)**

#### **a) Perturbations au niveau de l'acide lipotéichoïque**

L'acide lipotéichoïque (LTA) est un polymère anionique de polyglycérol-phosphate ancré à la membrane bactérienne par un glycolipide ancreur. Ce polymère peut subir des modifications une fois exposé à la surface de la bactérie, tel que l'ajout de résidus de D-alanine, qui, en apportant des fonctions amines, neutralisent les fonctions acides liées aux groupements

phosphates. Le taux de D-alanylation est donc un moyen pour la cellule de contrôler la charge de l'enveloppe. Cette charge est impliquée dans de nombreux aspects de l'interaction de la cellule avec son environnement, et notamment dans la sensibilité aux peptides cationiques. Dans ce cadre, j'ai analysé précisément la réponse transcriptomique de la cellule à l'absence de D-alanylation du LTA, et la régulation de l'expression de l'opéron *dltABCD* (Article 1). J'ai aussi étudié l'effet de l'addition de colistine en condition sub-inhibitrice pour comparer cette réponse à l'effet de la mutation *dltA*, et pour déterminer si l'addition de peptides cationiques avait un effet sur l'expression de l'opéron *dltABCD*. Enfin j'ai tenté d'obtenir un mutant touchant directement la synthèse de l'acide lipotéichoïque.

### **(1) Régulations en relation avec le taux de D-alanylation du LTA (publication n°1)**

Le locus *dlt* est constitué, chez *S. agalactiae*, de l'opéron *dltABCD*, précédé par les gènes *dltRS*, codant pour un système à deux composants. Il avait été montré que l'expression de ce locus dépendait de deux promoteurs, tous les deux activés en absence de D-alanylation [141]. Le premier, *PdltR*, en amont de *dltR*, est sous le contrôle de DltR. Le second, *PdltA*, en amont de *dltA*, est indépendant de DltR [141]. Par ailleurs, l'analyse du phénotype d'un mutant du gène *dltA*, ne modifiant plus le LTA, avait mis en évidence l'effet pléiotropique de cette mutation, touchant la virulence et la sensibilité aux peptides cationiques [100]. Afin de comprendre la réponse de *S. agalactiae* à une absence de D-alanylation du LTA, j'ai dans un premier temps, entrepris une analyse transcriptomique du mutant *dltA* comparé à la souche sauvage NEM316. Cette étude a montré que dans les conditions testées, malgré l'effet pléiotrope de cette mutation, seulement 120 gènes avaient une expression modifiée, cette modification correspondant en majorité à une sur-expression dans le mutant *dltA* (113 gènes). Une partie de la réponse de la bactérie implique l'induction de nombreux gènes d'adaptation au stress. L'absence de D-alanylation entraîne aussi l'induction de trois systèmes à deux composants: DltRS, qui était attendu selon les données publiées, CiaRH et RelRS. L'induction de ces deux autres systèmes était intéressante, dans le contexte de l'analyse de systèmes de régulations, répondant potentiellement à des modifications de charge de l'enveloppe. CiaRH correspond avec les systèmes WalKR (*gbs0741-42*) et LiaSR (*gbs0309-10*) aux trois TCS retrouvés chez tous les streptocoques (Tableau S1). Il a été montré que CiaRH est impliqué dans l'adaptation au stress acide chez *S. mutans*. Il faut remarquer que l'acidité est un paramètre important de l'environnement des bactéries lactiques. Il existe un lien établi entre diminution du pH et

augmentation du taux de D-alanylation du LTA chez *S. aureus* et *B. subtilis* [331]. L'implication du système LiaSR a été démontrée dans la régulation de la D-alanylation du LTA chez *S. gordonii* [332] et le système WalKR est essentiel, et régule le métabolisme de la paroi [333]. Ces deux TCS sont exprimés de manière constitutive chez *S. agalactiae*, et ne sont différenciellement exprimés dans aucune des conditions testées parmi les transcriptomes analysés. Le troisième TCS dont l'expression est induite par l'absence de D-alanylation, *relRS*, est co-transcrit avec le gène *relP* qui code pour une ppGpp synthase. Le rôle physiologique de ce TCS restant encore mal connu malgré son étude dans d'autres streptocoques [227, 334].

Les systèmes à deux composants sont fréquemment autorégulés; ainsi, leurs surexpressions dans le mutant *dltA* pouvaient suggérer que ces TCS étaient potentiellement impliqués dans la réponse de la bactérie à l'absence de D-alanylation. Afin de disséquer le rôle de ces trois TCS, *DltRS*, *RelRS* et *CiaRH*, j'ai réalisé des analyses transcriptomiques sur les souches mutantes pour *dltR*, *ciaRH*, *relRS*, ainsi que sur les doubles mutants *dltRdltA*, *ciaRHdltA*, *relRSCiaRH*, les souches me servant de références étant les souches NEM316 et NEM316  $\Delta dltA$ . Cette analyse m'a permis de montrer que, 62% de la réponse à l'absence de D-alanylation dépendait directement ou indirectement de ces trois TCSs, mettant en évidence leurs implications dans la réponse de *S. agalactiae* à des perturbations du niveau de charge, à la surface de la bactérie, résultant de l'absence de D-alanylation.

Cette analyse a également été le point de départ pour une analyse en profondeur du régulon *CiaR*, et m'a permis d'identifier 18 promoteurs, probablement sous le contrôle direct de ce régulateur, dont quatre contrôlent la synthèse de petits ARN non codants. L'analyse des protéines codées par les autres gènes du régulon montre qu'elles peuvent avoir un rôle au niveau de l'adaptation dans des conditions acides: protéase membranaire, chaperone, ainsi que des protéines membranaires ou secrétées de fonctions inconnus. Tous ces gènes, sauf *gbs1415*, possèdent un site potentiel de fixation de *CiaR*, suggérant une régulation directe au niveau des séquences promotrices. De manière intéressante, mis à part les gènes codant pour les petits ARN et *htrA*, le régulon *CiaRH* est assez différent de celui caractérisé chez d'autres streptocoques, montrant la spécificité de la réponse de *S. agalactiae*. En lien plus direct avec ce travail, nous avons aussi montré que l'opéron *relPRS* était contrôlé directement par *CiaRH*, ainsi que le promoteur *PdltA*.

Enfin, j'ai pu montrer que la transcription du locus *dlt* était plus complexe qu'initialement décrite, et dépendait de trois promoteurs. Le premier, le plus en amont,  $P_{0dltR}$  constitutivement exprimé, et les deux autres  $P_{dltR}$  et  $P_{dltA}$  fortement induits par l'absence de D-alanylation, et régulés respectivement par DltRS et CiaRH. L'existence du promoteur en amont de  $P_{dltR}$  ( $P_{0dltR}$ ) a été suggérée par des expériences d'identification des sites d'initiation de transcription par RNAseq, réalisées au laboratoire par Isabelle Rosinski-Chupin en collaboration avec Odile Sismeiro (PF2, Génopôle). L'existence de ce promoteur a par la suite également été confirmée par l'analyse de l'expression d'une fusion avec le gène *lacZ* d'*E. coli*. J'ai pu confirmer, par mutagenèse dirigée, les sites de fixation prédits pour DltR et CiaR au niveau des séquences -35 des promoteurs respectivement  $P_{dltR}$  et  $P_{dltA}$ , dont l'altération de ces séquences abolit l'activation par DltR et par CiaR. La double régulation de l'expression de l'opéron *dltA-D* par DltR et CiaR permet de proposer différents modèles pour expliquer les phénotypes de sensibilité à la colistine et à l'acidité des différentes souches mutantes. La souche  $\Delta dltA$  est très sensible à ces deux paramètres tout comme le double mutant  $\Delta dltR-\Delta ciaRH$ , alors que les souches simple-mutantes  $\Delta dltR$  et  $\Delta ciaRH$  ont un phénotype intermédiaire. Les données du transcriptome du mutant  $\Delta ciaRH$ , par rapport au sauvage, indiquent que, dans les conditions testées, la mutation de ce système a peu d'effet sur l'expression des gènes *dltABCD*. Le phénotype intermédiaire du mutant CiaRH laisse supposer, soit que les différents transcrits issus de  $P_{dltR}$  et  $P_{dltA}$  ne sont pas traduits avec la même efficacité et que les ARN initiés à partir de  $P_{dltA}$  sont traduits plus efficacement; soit que l'effet de l'absence de CiaRH sur le phénotype de sensibilité à la colistine et à l'acidité est indépendant de la D-alanylation du LTA, et résulte de la diminution d'expressions d'autres protéines sous le contrôle de CiaR. Dans cette alternative, cet effet se cumulerait avec l'effet de l'absence de DltR sur l'expression de l'opéron *dltABCD*, entraînant la forte sensibilité du double mutant à la colistine et au pH acide.

L'ensemble de nos résultats suggère que si l'activation transcriptionnelle des systèmes DltRS et CiaRH est probablement une conséquence directe de l'absence de D-alanylation, l'induction de *relRS* est due à l'activation du système CiaRH. Nos résultats montrent que le rôle de RelRS dans la réponse à l'absence de D-alanylation semble limité. Néanmoins, la conservation de RelRS au sein du genre streptocoque indique que ce système pourrait jouer un rôle important au niveau des réseaux de régulation chez *S. agalactiae* et les autres streptocoques. Le rôle central de RelRS, également appelé SptRS (pour "saliva persistence") chez *S. pyogenes* a été suggéré par deux études. La première montre que chez *S. pyogenes* l'expression de *sptRS*

est augmentée en présence de salive et que ce TCS contribue à la survie de *S. pyogenes* dans la salive [227]. La seconde montre que la délétion du régulateur conduit à l'expression différentielle de 342 gènes, que ce mutant est hyper virulent dans un modèle d'infection cutanée chez la souris, avec une apparition plus rapide d'abcès de tailles plus importantes [334]. De même, l'expression de *sptRS* est fortement induite dans des souches responsables de pharyngites isolées chez six patients, suggérant l'importance de ce système *in vivo* pour la pathogénicité de *S. pyogenes* [227]. Chez *S. agalactiae*, la délétion de *relRS* entraîne l'expression différentielle de 123 gènes. Parmi ces gènes, aucun facteur de virulence connu n'est retrouvé, cependant trois d'entre eux codent pour des protéines de surface avec des domaines CHAP et LrgAB (*gbs1805*, *gbs1214* et *gbs1215*), potentiellement impliqués dans l'hydrolyse du peptidoglycane. Il sera intéressant de tester la virulence du mutant *relRS* dans des modèles d'infections murins, afin de déterminer si, comme chez *S. pyogenes*, RelRS pourrait avoir un rôle dans le contrôle de la virulence chez *S. agalactiae*.

Le résultat le plus important de ce travail est la mise en évidence que le processus de D-alanylation, chez *S. agalactiae*, est contrôlé de manière complexe par deux boucles de rétroaction faisant intervenir deux systèmes de régulation. Le niveau de D-alanylation module l'activité de DltR et de CiaR et l'expression de l'opéron *dlt* est en retour régulée par ces deux régulateurs. A ce jour, cette double régulation semble unique à *S. agalactiae* parmi les streptocoques. Nous avons proposé que l'expression de l'opéron *dltA-D* pourrait être régulée par DltR chez *S. infantarius*. Au contraire, l'opéron *dltA-D* est sous le contrôle direct ou indirect de CiaR chez *S. pneumoniae* [208]. Dans les deux cas, la régulation par l'autre TCS ne semble pas pouvoir intervenir, l'opéron *dltRS* n'étant pas présent chez *S. pneumoniae* et le motif de régulation par CiaR manquant dans la région promotrice de *dltA* chez *S. infantarius*. Cette double régulation de l'opéron *dlt* chez *S. agalactiae* permet un ajustement fin du taux de D-alanylation et de la charge nette à la surface de la bactérie. Nos résultats montrent également que les deux systèmes de régulation interviennent conjointement dans la sensibilité aux peptides cationiques, confirmant l'importance du taux de D-alanylation dans la sensibilité de *S. agalactiae* aux peptides antimicrobiens.

## **Publication n°1**

---

**D-alanylation of lipoteichoic acid in *Streptococcus agalactiae* is under the control of a dual feed back regulatory loop involving DltRS and CiaRH**

Journal:	<i>Molecular Microbiology</i>
Manuscript ID:	Draft
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Da Cunha, Violette; Institut Pasteur, Dept. Génomes et Génétique Caliot, Elise; Institut Pasteur, Dept. Microbiologie Rosinski-Chupin, Isabelle; Institut Pasteur, Dept. Génomes et Génétique Couvé, Elisabeth; Institut Pasteur, Dept. Génomes et Génétique Proux, Caroline; Institut Pasteur, Dept. Génomes et Génétique Rusniok, Christophe; Institut Pasteur, Dept. Génomes et Génétique Poyart, Claire; Université Paris Descartes, Institut Cochin Trieu-Cuot, Patrick; Institut Pasteur, Dept. Microbiologie Glaser, Philippe; Institut Pasteur, Dept. Génomes et Génétique
Key Words:	Bacterial cell envelope, Gene regulation, transcriptome, Virulence, cationic antimicrobial peptide



1 **D-alanylation of lipoteichoic acid in *Streptococcus agalactiae* is under the control of a dual**  
2 **feed back regulatory loop involving DltRS and CiaRH**

3

4

5 Running title: Regulation of D-alanylation in *S. agalactiae*

6

7 Violette Da Cunha<sup>1,2</sup>, Elise Caliot<sup>3,4</sup>, Isabelle Rosinski-Chupin<sup>1,2</sup>, Elisabeth Couvé<sup>1,2</sup>, Caroline  
8 Proux<sup>5</sup>, Christophe Rusniok<sup>2,6</sup>, Claire Poyart<sup>3,4,7,8</sup>, Patrick Trieu-Cuot<sup>3,4,\*</sup>, Philippe Glaser<sup>1,2,\*</sup>

9

<sup>1</sup>Institut Pasteur, Laboratoire Evolution et Génomique Bactériennes, 28 rue du Dr Roux 75724 Paris; <sup>2</sup>CNRS URA 2171 28 rue du Dr Roux 75724 Paris; <sup>3</sup>Institut Pasteur, Unité de Biologie des Bactéries Pathogènes à Gram Positif, 28 rue du Dr Roux 75724 Paris; <sup>4</sup>CNRS URA 2172, 28 rue du Dr Roux 75724 Paris; <sup>5</sup>Institut Pasteur, Genopole Plate-Forme Transcriptome et Epigénome, 28 rue du Dr Roux 75724 Paris; <sup>6</sup>Institut Pasteur Unité de Biologie des Bactéries Intracellulaires, 28 rue du Dr Roux 75724 Paris; <sup>7</sup>Institut Cochin, Université Paris Descartes, Sorbonne Paris Cité, CNRS (UMR 8104), Paris, France, <sup>8</sup>INSERM, U1016, Paris, France.

10

11 \*To whom correspondence should be addressed

12 Philippe Glaser

13 Laboratoire Evolution et Génomique Bactériennes, Institut Pasteur, 28 rue du Dr Roux, 75724

14 Paris cedex 15 France. Tel: 33 1 45 68 89 96, Fax: 33 1 45 68 87 86

15 email: pglaser@pasteur.fr

16 and Patrick Trieu-Cuot

17 Unité de Biologie des Bactéries Pathogènes à Gram Positif, Institut Pasteur, 28 rue du Dr

18 Roux, 75724 Paris cedex 15 France. Tel: 33 1 45 68 95 92

19 email: ptrieu@pasteur.fr

20

20 **SUMMARY**

21

22 D-alanylation of teichoic acids (TA) of the Gram-positive cell wall, catalyzed by the  
23 *dltABCD* operon, modulates the bacterial surface charge. In *Streptococcus agalactiae*  
24 inactivation of the *dlt* operon has a pleiotropic effect, including decreased virulence. To gain a  
25 deeper understanding of its role, we analysed *S. agalactiae* transcriptional response to D-  
26 alanyl deprivation and identified 113 genes induced, including those encoding the two-  
27 component systems (TCSs) DltRS, CiaRH and RelRS, and 7 genes repressed. Mutational and  
28 transcriptional analysis of these TCSs showed that they regulate directly or indirectly 62% of  
29 the gene expression changes observed. Bioinformatics analysis of promoter regions identified  
30 binding sites for DltR and CiaR and defined precisely 19 genes as the CiaRH regulon. Two  
31 promoters, *PdltR* and *PdltA* control the *dltABCD* operon. In D-Ala-deficient TA mutants,  
32 *PdltR* is activated by DltR and we show here that *PdltA* is activated by CiaRH. Therefore, in  
33 *S. agalactiae*, the homeostasis of the bacterial cell surface net charge involves a dual feed  
34 back regulatory loop as DltR and CiaR respond to a low level of D-alanylation by inducing  
35 the transcription of the *dlt* operon. This regulatory loop links LTA D-alanylation to the  
36 adaptation to acidic conditions mediated by CiaRH.

37

37 **INTRODUCTION**

38

39 *Streptococcus agalactiae*, also known as Group B Streptococcus (GBS), was initially  
40 described as responsible for bovine mastitis colonizing the udder canals (Keefe, 1997) but is  
41 now recognized as a leading cause of infections in neonates causing pneumonia, septicemia,  
42 and meningitis (Dermer *et al.*, 2004). However, this bacterium is primarily a commensal of  
43 the human digestive and genitourinary tracts and newborns are usually colonized during birth  
44 by GBS strains present in the vaginal flora of their mother. In addition, GBS is an increasing  
45 cause of various infections including urinary tract infections, endocarditis, osteomyelitis,  
46 septicaemia and meningitis in elderly and persons with underlying diseases (Phares *et al.*,  
47 2008). Thus, GBS have the capacity to colonize and to adapt to a broad range of  
48 environments, tissues and hosts.

49 The cell wall of low GC Gram-positive bacteria is a complex macromolecular  
50 assembly of cross-linked peptidoglycan, polyanionic teichoic acids (TAs), and surface  
51 proteins essential for cellular integrity and shape, and the resistance to external stresses and  
52 aggressions (Silhavy *et al.*, 2010). TAs are composed of wall teichoic acid (WTA) covalently  
53 linked to the peptidoglycan and lipoteichoic acid (LTA) anchored in the membrane *via* a  
54 glycolipid moiety. WTA and LTA are assembled *via* different pathways and, interestingly,  
55 streptococci, enterococci and lactococci apparently lack the genes coding the essential  
56 enzymes for glycerophosphate WTA synthesis that are present in bacilli, lactobacilli, listeria,  
57 and staphylococci (Weidenmaier & Peschel, 2008) and our data not shown.

58 The external repetitive unit of LTAs and of many WTAs consists of polyglycerol-  
59 phosphate that forms a chain extending into the cell wall. These polyanionic polymers may be  
60 neutralized by the addition of protonated D-alanyl ester residues that provide counterions for  
61 determining the net anionic charge of the TAs. Incorporation of D-Ala residues into LTAs and

62 WTAs necessitates the activity of four proteins encoded by the *dltABCD* operon (Perego *et*  
63 *al.*, 1995). Inactivation of any gene within this operon results in the complete absence of D-  
64 Ala ester in TAs. These D-Ala-deficient TA mutants exhibit a variety of phenotypic changes  
65 that were attributed to the net increase of the polyanionic surface charge. In particular, these  
66 mutants were able to more efficiently bind cationic compounds including detergents  
67 (Debabov *et al.*, 2000), autolysins (Wecke *et al.*, 1997), adhesins (Clemans *et al.*, 1999), and  
68 antimicrobial peptides (Peschel *et al.*, 1999, Poyart *et al.*, 2003, Abachin *et al.*, 2002, Kristian  
69 *et al.*, 2005, Kovacs *et al.*, 2006). Defects in D-alanyl-TA synthesis can also lead to acid  
70 sensitivity, as demonstrated for *Streptococcus mutans* (Boyd *et al.*, 2000), presentation of  
71 secreted proteins at the cell surface, as shown for *B. subtilis* (Hyyrylainen *et al.*, 2000) or cell  
72 separation in *Lactobacillus plantarum* (Palumbo *et al.*, 2006). D-Ala esterification of TAs is  
73 essential for virulence of *Bacillus cereus* (Abi Khattar *et al.*, 2009), *Streptococcus agalactiae*  
74 (Poyart *et al.*, 2003), *Streptococcus pyogenes* (Cox *et al.*, 2009) and *Staphylococcus aureus*  
75 (Collins *et al.*, 2002, Weidenmaier *et al.*, 2004) and the intracellular pathogen *Listeria*  
76 *monocytogenes* (Abachin *et al.*, 2002). Thus, a tight regulation of this process is important to  
77 control the cell envelope homeostasis.

78 In *S. agalactiae*, the *dltABCD* genes are preceded by two genes, *dltR* and *dltS*,  
79 encoding a two-component system (TCS) (Poyart *et al.*, 2001). To date, among Firmicutes,  
80 this organization has only been described for *S. agalactiae* and *Streptococcus infantarius* (Fig.  
81 1). The regulatory protein DltR belongs to the OmpR family and the sensor protein DltS  
82 contains two transmembrane domains suggesting that it is a membrane associated protein with  
83 an extracellular domain involved in the sensing of environmental stimuli. The expression of  
84 the GBS *dltABCD* genes was reported to be controlled by two promoters designated *PdltR* and  
85 *PdltA* (Fig. 1). *PdltR* is located immediately upstream *dltR* and directs the synthesis of a  
86 transcript extending from *dltR* to *dltD*. This promoter is activated by DltR in mutants deficient

87 in TA D-Ala ester substitution. *PdltA* is located upstream *dltA* in the 3' extremity of *dltS* and  
88 directs synthesis of a transcript encompassing *dltABCD*. This promoter is activated in D-Ala  
89 deficient LTA mutants independently of DltR by a yet unknown regulatory mechanism  
90 explaining why the level of D-alanylation is only marginally reduced in a *dltR* mutant (Poyart  
91 et al., 2001).

92 In order to decipher the regulatory network involved in sensing the D-alanylation  
93 status of the cell and regulating the expression of the GBS *dlt* operon, a genome-wide  
94 transcriptional analysis of WT and D-Ala deficient LTA mutant strains was carried out. D-  
95 alanine ester deprivation of LTAs induces the increased expression of three TCSs: *dltRS*,  
96 *ciaRH*, and *relRS*. Further analyzes of the transcriptional effect of the inactivation of these  
97 different regulatory systems, either in the presence or absence (*dltA* mutant) of LTA D-  
98 alanylation, allowed to determine their specific involvement in the response to the absence of  
99 D-alanylation of LTA and in the expression of the *dlt* operon. The combined results allow to  
100 propose a model where the control of LTA D-alanylation is under the dual control of DltRS  
101 and CiaRH and involves two feedback control loops that ensure a fine tuning of the net charge  
102 of the bacterial cell envelope.

103

103 **RESULTS**

104

105 **The transcriptional response of *S. agalactiae* to the absence of LTA D-alanylation**

106 The deletion of the *dltA* gene ( $\Delta dltA$ ) leads to the complete absence of D-Ala ester in GBS  
107 LTAs (Poyart et al., 2001). In order to define the physiological response to this modification  
108 of the cell envelope, we have compared the transcription profiles of strain NEM316 and its  
109  $\Delta dltA$  derivative. Bacteria grown in TH broth were harvested in early exponential phase  
110 before medium acidification, *i.e.* in conditions where the growth rate of the  $\Delta dltA$  strain was  
111 similar to that of the NEM316 strain (data not shown). Although the absence of D-alanylation  
112 has a pleiotropic effect on the cell physiology (Poyart et al., 2001, Poyart et al., 2003), only  
113 120 genes (5.6%) were differentially expressed ( $\geq 2$ -fold change) in the  $\Delta dltA$  strain of which  
114 113 genes were upregulated but only seven downregulated (Table 1).

115 Analysis of the function of these genes showed a clear induction of the general stress  
116 response, as the *hrcA-grpE-dnaK-dnaJ* and *groES-groEL* operons were highly upregulated as  
117 well as several other stress associated genes like *clpP* (*gbs1634*), *clpE* (*gbs0535*), *clpL*  
118 (*gbs1376*), the foldase gene *prsA* and the arginine deiminase pathway (*gbs2084-2085*). We  
119 also observed the induction of 5 genes encoding ribosomal proteins (L28, S15, S20, S21, and  
120 L34) and two encoding ribosome associated proteins (Gbs0326, Gbs0581). The LTA  
121 backbone is composed of glycerol and phosphate and interestingly a set of genes involved in  
122 glycerol or phosphate metabolism was 2 to 3-fold induced in absence of D-alanylation:  
123 *gbs1695* encoding a dihydroxyacetone kinase, *gbs1697* encoding a glycerol uptake protein,  
124 and *gbs1950-1953* encoding a phosphate transporter. Furthermore, we observed a 2.5-fold  
125 induction of the extracellular nuclease Gbs0661 that, by degrading nucleic acids, might  
126 contribute to the release of phosphate. The negative charge of the envelope leads probably to  
127 a reduction of the local concentration of the anionic phosphate ions that may activate the

128 expression of these genes. Although the absence of D-alanyl residues on LTA led to major  
129 alterations in the cell envelope, it did not cause modifications in the expression of genes  
130 encoding other surface components as the polysaccharidic capsule and antigen B operons or  
131 those encoding surface proteins and pili. The only exceptions were *fbsA*, coding for a  
132 fibrinogen binding protein, and *cspA* (*gbs2008*), coding for the cell surface serine protease  
133 CspA, that were 2-fold induced and 2-fold repressed in the  $\Delta dltA$  mutant, respectively.

134 Several regulatory genes displayed altered expression in the D-alanylation deficient  
135 mutant. It is likely that they contribute to the observed transcriptional response. Eight  
136 regulators were induced in the  $\Delta dltA$  mutant strain; all but one (*gbs1972*) belonging to the  
137 GBS core genome. Among the five stand-alone regulators, two (HrcA and SpxA (Gbs2046))  
138 have been shown to be involved in the stress response in other *Streptococci* (Jayaraman *et al.*,  
139 1997, Kajfasz *et al.*, 2010). The three remaining stand-alone regulators (Gbs0168 belonging  
140 to the C1/Cro family, Gbs1402 to the MarR family, and Gbs1972) are yet uncharacterized.  
141 Most interestingly, we also identified three two-component regulatory systems (TCS) that  
142 were induced in the mutant strain: DltRS, as expected from our previous work demonstrating  
143 its role in modulating the expression of the *dlt* operon (Poyart *et al.*, 2001); CiaRH (Gbs1019-  
144 1020) and RelRS (Gbs1397-1398) which were not previously reported to be implicated in this  
145 regulatory network. Like DltR, CiaR and RelR belong to the OmpR family of regulators that  
146 are frequently involved in sensing changes in environmental conditions (Itou & Tanaka,  
147 2001). Noticeably, deletion of the *S. agalactiae* *ciaR* gene led to a decrease in intracellular  
148 survival in different mouse and human immune cell types which was attributed to an  
149 increased sensitivity to cationic antimicrobial peptides by an uncharacterized mechanism  
150 (Quach *et al.*, 2009). As in *Streptococcus mutans*, the *relRS* homologs are located downstream  
151 *relP* which encodes a ppGpp synthase (Seaton *et al.*, 2010), an enzyme directing synthesis of  
152 the alarmone ppGpp involved in the response to different nutritional stresses in Gram-positive

153 and Gram-negative bacteria (Wolz *et al.*, 2010). The observation of the induction of these  
154 three TCSs prompted us to define their respective contribution to the response of the bacteria  
155 to the absence of LTA D-alanylation and to analyze whether these three TCSs interact.

156

### 157 **DltRS, CiaRH, and RelRS TCSs contribute to the transcriptional response to the** 158 **absence of LTA D-alanylation**

159 To characterize the role of the TCSs DltRS, CiaRH and RelRS in transcriptional  
160 regulation related to D-Ala ester deprivation, we generated single ( $\Delta dltR$ ,  $\Delta ciaRH$ , and  
161  $\Delta relRS$ ) and double mutations ( $\Delta dltR\Delta dltA$ ,  $\Delta ciaRH\Delta dltA$ , and  $\Delta relRS\Delta dltA$ ) in NEM316.  
162 The transcriptional program of single mutants were compared to strain NEM316 and the  
163 double mutants were compared to the  $\Delta dltA$  strain to identify genes regulated directly or  
164 indirectly by each TCS in the absence of D-alanylation. We focused our analysis on the 120  
165 genes that we identified as altered in the  $\Delta dltA$  strain (Table 1).

166 To define the role of DltR in response to changes in LTA D-alanylation, we compared  
167 the transcriptional program of *S. agalactiae* in four different genetic backgrounds:  $\Delta dltR/WT$ ,  
168  $\Delta dltA/WT$ ,  $\Delta dltR\Delta dltA /WT$ , and  $\Delta dltR\Delta dltA /\Delta dltA$ . The comparison of the transcriptome of  
169 the  $\Delta dltR$  and the WT strains revealed 244 (11.4% of NEM316 genes) differentially expressed  
170 genes (Table S1) of which 75 were downregulated and 169 upregulated in the  $\Delta dltR$  strain.  
171 Thus, although no effect on the bacterial growth was detectable in TH medium, the  
172 inactivation of *dltR* leads to profound changes in the transcription profile. Many of these  
173 changes may be indirectly mediated by other regulatory systems as the expression of nine  
174 regulators including two TCS RgfAC (Gbs1944-43) and VncR (Gbs0597-98) and the orphan  
175 response regulator Gbs0685 was modified. Comparative analyses of the four transcriptome  
176 experiments confirmed that the activation of the expression of *dltR* and *dltS* in the absence of  
177 D-alanylation was DltR-dependent (Poyart *et al.*, 2001) as they were inversely regulated in the



178  $\Delta dltA$ /WT and  $\Delta dltR\Delta dltA/\Delta dltA$  comparisons and not differentially expressed in the  
179  $\Delta dltR\Delta dltA$ /WT comparison (Table S1). We identified 29 genes showing a transcription  
180 pattern similar to *dltR* and *dltS*, suggesting that they are regulated directly or indirectly by  
181 DltR in the response to D-alanyl ester depletion of LTAs (Table S1). These genes include the  
182 10 genes belonging to the stress response previously described, including the seven genes  
183 regulated by HrcA (Table 1). They include also those involved in glycerol metabolism and  
184 phosphate homeostasis, which suggests a link between DltR regulation and LTA biosynthesis.  
185 In addition, four genes corresponding to metabolic adaptation (*gbs0189*, *gbs0190*, *gbs1116*  
186 *gbs1683*) showed an opposite expression pattern. They are involved in trehalose metabolism  
187 and in permeation of amino acids and xanthine. In contrast, the *ciaRH* and *relPRS* operons  
188 were among the genes differentially expressed both in the  $\Delta dltA$ /WT and  $\Delta dltR\Delta dltA$ /WT  
189 comparison, but not in the  $\Delta dltR\Delta dltA/\Delta dltA$ . These results suggest that these TCSs are  
190 upregulated in the absence of D-alanylation independently of DltR.

191 To further our knowledge on the role of these two TCSs in this regulatory network, we  
192 undertook the expression analysis of the *ciaRH* deletion in different genetic backgrounds. Our  
193 transcriptome analysis identified 20 downregulated and 64 upregulated genes in the  $\Delta ciaRH$   
194 strain as compared to the WT strain (Table S2), a number much higher than previously  
195 reported for the gene expression of a  $\Delta ciaR$  mutant in strain COH1 where only six genes were  
196 reported as downregulated (SAN\_1101 (*ciaR*), SAN\_2180, SAN\_0039 (duplicated as  
197 SAN\_0058 and SAN\_0059 in the COH1 strain annotations), SAN\_2168, SAN\_2294 and  
198 SAN\_0238) and one as upregulated (SAN\_2409) in the *ciaR* mutant (Quach et al., 2009). The  
199 here identified genes include the orthologs of the genes identified by Quach et al. except  
200 SAN\_2294, which is missing in strain NEM316. To further define the role of the CiaRH TCS  
201 we then investigated the impact of the deletion of this system in absence of D-alanylation by  
202 comparing the gene expression program of the *ciaRH dltA* double mutant to the *dltA* single

203 mutant ( $\Delta$ *ciaRHdltA*/ $\Delta$ *dltA*). In the  $\Delta$ *dltA* background, *ciaRH* inactivation led to the  
204 downregulation of 60 genes and the upregulation of 110 genes (Table S2). Interestingly, the  
205 absence of D-alanylation led to an increased induction or repression of 133 of the 170 genes  
206 affected by the inactivation of CiaRH showing the impact of D-alanylation on the regulations  
207 mediated by CiaR.

208         Among these 170 genes, 41 belong to the 120 genes regulated by the absence of D-  
209 alanylation ( $\Delta$ *dltA* mutant) (Table 1). Twelve of these 41 genes are induced in the absence of  
210 D-alanylation but are downregulated in the  $\Delta$ *ciaRH* $\Delta$ *dltA* strain (Table 1 and Fig. 2A)  
211 suggesting that their increased expression in the  $\Delta$ *dltA* mutant is due to CiaR activation.  
212 Interestingly, the *relPRS* operon belongs to this class indicating that the *relRS* TCS is  
213 regulated by CiaR and therefore not autoregulated by the absence of LTA D-alanylation.  
214 Among the remaining genes, five were also repressed in the  $\Delta$ *ciaR* mutant compared to the  
215 WT strain (*gbs0030*, *gbs0196*, *gbs1995*, *gbs2003*, *gbs2133*). Finally, 24 genes were induced  
216 in the absence of D-alanylation and their expression was further increased by the deletion of  
217 *ciaRH* in the  $\Delta$ *dltA* background. Nineteen out of these 24 genes, including those of the stress  
218 regulon controlled by HrcA, showed an opposite effect of the mutation of *ciaRH* and *dltR*.  
219 These results indicate that the stress due to D-alanyl ester deprivation is alleviated by  
220 inactivation of *dltR* whereas it is increased by the deletion of *ciaRH*.

221         To also determine the contribution of RelRS to the regulations involved in the  
222 response to the absence of D-alanylation, we analyzed the transcriptional impact of the  
223 deletion of the *relRS* and  $\Delta$ *relRS* $\Delta$ *dltA* genes compared to the WT and the  $\Delta$ *dltA* strains  
224 respectively (Table S3). Among the 120 genes differentially expressed in the  $\Delta$ *dltA* strain,  
225 only nine were affected by the deletion of *relRS* in the double mutant  $\Delta$ *relRS* $\Delta$ *dltA* (seven  
226 were upregulated and two downregulated). The expression of seven of these genes is also  
227 affected by mutating either *ciaRH* or *dltR* (Table 1). Only the operon *gbs2083-2085* encoding

228 a carbamate kinase and an ornithine carbamoyltransferase that was strongly induced in the  
229  $\Delta dltA$  strain was also induced specifically in the  $\Delta relRS$  mutant. Thus, the contribution of the  
230 RelRS TCS to the response of *S. agalactiae* to D-alanyl ester deprivation is probably minor.

231 Lastly, the expression of 45 genes differentially expressed in the *dltA* mutant  
232 compared to NEM316 was not affected by the deletion of any of the three TCSs. The  
233 modification of their expression might thus be mediated by one of the five stand-alone  
234 regulators whose expression was induced in the *dltA* mutant. Interestingly, except *hrcA*, four  
235 of these five regulators were induced independently of DltRS, CiaRH, and RelRS.

236

### 237 **Prediction of CiaR and DltR DNA binding sites**

238 The DNA binding site of CiaR has been experimentally identified in *S. pneumoniae*  
239 and in *S. mutans*. The consensus sequence is conserved in these two species and corresponds  
240 to a repetition of the hexanucleotide motif TTAAAG separated by 5 nts located 10 nts  
241 upstream the -10 box of the promoter (Halfmann *et al.*, 2007, Wu *et al.*, 2010, Marx *et al.*,  
242 2010). This architecture is characteristic of a Class II activation mechanism where the  
243 activator binds to a target that overlaps the promoter -35 element and contacts domain 4 of  
244 the RNA polymerase  $\sigma$  subunit (Browning & Busby, 2004). We thus searched for such motifs  
245 in the promoter sequences of the 122 transcriptional units identified in this study (Table S2) as  
246 being regulated directly or indirectly by CiaR in WT and  $\Delta dltA$  genetic backgrounds. This  
247 revealed the presence of a similar repeated hexanucleotide motif at the expected position in  
248 eight promoters (Fig. 2A). Two of these promoters control the expression of the *ciaRH* and  
249 *relPRS loci*. Among the six other genes two encode secreted peptidases (Gbs0030 and HtrA)  
250 and two membrane proteins with putative oxidoreductase (Gbs0196) and phosphatase  
251 (Gbs1995) activity. A single gene (*gbs1415*) predicted on the basis of the transcriptional data

252 to belong to the CiaR regulon does not present this hexanucleotide motif in its promoter  
253 suggesting an indirect regulation by CiaR.

254 In an attempt to determine if the genes identified as differentially expressed by the  
255 transcriptome analysis define the CiaR regulon, we systematically searched for this putative  
256 CiaR binding sequence upstream all -10 promoter sequences as determined from RNAseq  
257 analysis of transcription start sites (TSS) (data not shown). In addition to the eight sequences  
258 described above, we identified ten putative CiaR binding sites (Fig. 2A). Interestingly, a CiaR  
259 binding sequence was identified upstream the *PdltA* promoter although deletion of this  
260 regulator had no apparent effect on the transcription of the *dlt* operon as quantified by DNA  
261 array hybridization (Table 1 and S2). In this transcriptome experiment, the CiaR-dependent  
262 *PdltA* transcription of the *dlt* operon may be masked by the activity of the upstream promoter  
263 *PdltR* (Fig. 1). A typical CiaR regulated promoter was detected upstream the *ftsA* gene  
264 although *ciaR* deletion had no effect on *ftsA* expression in our transcriptome experiments.  
265 This might be due to the presence of a CiaR-independent stronger promoter located between  
266 *PftsA* and the initiation codon as determined by RNAseq (data not shown). Four of the 18  
267 putative CiaR binding sites are located upstream promoter sequences that might be  
268 responsible for the expression of small non-coding RNAs that were recently predicted in an *in*  
269 *silico* analysis (Marx et al., 2010). As the genes encoding these four csRNA were not  
270 represented on the DNA-array, their expression was tested by quantitative RT-PCR in  
271 NEM316 WT,  $\Delta dltA$ ,  $\Delta ciaRH$ , and  $\Delta ciaRH\text{-}\Delta dltA$  mutant strains. Due to the small size and  
272 the sequence of these csRNAs, we were able to design primer pairs to study the transcription  
273 of only two of them, csRNA10 and csRNA11. qRT-PCR analysis revealed that they exhibited  
274 the same expression pattern as the genes activated by CiaR (Table 2). They were induced in  
275 the absence of D-alanylation and this activation was controlled by CiaR. The four other CiaR  
276 boxes identified are upstream *gbs0864*, *prsA* (*gb0827*), and *gbs0951*, which showed no

277 expression change in a *ciaRH* mutant although three of them, *prsA*, *gbs0951*, and *gbs1654*  
278 were induced in the  $\Delta dltA$  strain. Thus, these putative binding sites are either non functional  
279 or regulate the gene expression under specific conditions not tested here. The gene *prsA*  
280 encodes a secreted peptide prolyl isomerase involved in the folding of secreted proteins and  
281 *gbs0951* a conserved secreted protein. Therefore, in response to the external stimuli sensed by  
282 the CiaH sensor, CiaR mainly activates the expression of secreted and membrane bound  
283 proteins that contribute to the adaptation of *S. agalactiae*.

284 Until recently, *dltRS* genes were only described in *S. agalactiae*. However, by  
285 searching in the NCBI nr database, we identified a similar organization of the *dlt* locus in the  
286 *Streptococcus infantarius* subsp. *infantarius* strain ATCC BAA-102 as it contains *dltRS*  
287 orthologs upstream of the *dltA-D* genes. In an attempt to identify a potential DNA binding site  
288 of DltR, we compared the DNA region containing *PdltR* in *S. agalactiae* with that of  
289 *S. infantarius* (Fig. 3A). This region is highly conserved between the two species and both  
290 contain an almost identical duplication of 6 bases separated by 5 nts (TTTCAA <5nt>  
291 TTTCAA) eleven bps upstream the conserved -10 region. This sequence and its relative  
292 position to the transcription start site are reminiscent to that of the CiaR regulatory site (Fig.  
293 3B) indicating that they likely correspond to the DltR binding site. We did not identify a  
294 similar motif upstream the genes upregulated by DltR in the  $\Delta dltA$  background, indicating that  
295 these genes are likely regulated indirectly.

296

### 297 **The expression of the *dlt* operon is under the control of three promoters and two TCSs.**

298 Previous analyses have shown that the *dltABCD* genes are transcribed as a long  
299 transcript initiated at a promoter located upstream *dltR* (*PdltR*) and a shorter transcript  
300 initiated upstream *dltA* (*PdltA*) (Poyart et al., 2001). Both promoters were characterized by  
301 primer extension (Poyart et al., 2001). However, using high through put identification of TSS

302 in WT NEM316 (data not shown) we identified an additional promoter, located 170 bp  
303 upstream *PdltR* that we designated  $P_{0dltR}$  (Fig. 3A). This promoter was included in the 527 bp  
304 long DNA fragment fused upstream the *lacZ* gene used to analyse the expression of *PdltR*  
305 (Poyart et al., 2001) thereby precluding a complete understanding of their respective  
306 regulatory roles. Here, *PdltR* was thus cloned independently of  $P_{0dltR}$ , upstream the *lacZ* gene  
307 of pTCV-*lac* and its activity was measured in different genetic backgrounds (Table 3). In  
308 agreement with the transcriptome data and our previous report (Poyart et al., 2001), *PdltR* was  
309 induced 3-fold in the absence of D-alanylation, an induction that required DltR but not CiaR,  
310 or RelR. However, we observed that *PdltR* is devoid of significant basal activity in the  
311 absence of DltR. These results suggest that *PdltR* activity strictly depends on the presence of  
312 DltR for being expressed and being strongly activated in the absence of LTA D-alanylation  
313 (Table 3). To further characterize *PdltR*, two bases of the putative DltR binding motif were  
314 mutated and the activity of the resulting promoter,  $PdltR_{M0}$  was evaluated in NEM316 and  
315 relevant mutant derivatives. In this construction the modified adenine nucleotides at positions  
316 -25 and -26 in the second repeat were expected not to alter the -35 region of the promoter.  
317 However, as shown in Table 3, this modified promoter displayed a low basal activity in all  
318 strains tested and was no longer upregulated in the  $\Delta dltA$  genetic background. This result is  
319 consistent with our hypothesis that this directly repeated motif may be essential for binding of  
320 DltR and hence, for *PdltR* activity. Introduction of the same mutation of *PdltR* in the original  
321 construct containing both  $P_{0dltR}$  and *PdltR* allows to quantify the expression of  $P_{0dltR}$ . This  
322 analysis revealed that  $P_{0dltR}$  is a relatively weak promoter whose activity remained  
323 unchanged in all tested strains, indicating that it is not regulated by DltR or CiaR and does not  
324 respond to the D-alanine deprivation of LTA (Table 3).

325 We similarly determined the promoter activity of *PdltA* by using a *lacZ* fusion. *PdltA*  
326 is induced 7-fold in the absence of D-alanylation. The activation of this weak promoter in the

327 absence of LTA D-alanylation was not affected by the inactivation of *dltR* or *relRS* but was  
328 abolished by that of *ciaRH*, an observation consistent with our *in silico* detection of a CiaR  
329 binding site upstream *PdltA*.

330 DltR and CiaR display 62% similarity and their predicted binding sites also exhibit  
331 sequence identities (Fig. 3B). To test the specificity of these regulators for their respective  
332 promoters, we investigated the effect of swapping the putative regulator binding sites by site  
333 directed mutagenesis (Fig. 3B). We constructed the corresponding *lacZ* fusions and compared  
334 their expression in the WT strain and  $\Delta dltA$ ,  $\Delta ciaRH\Delta dltA$  and  $\Delta dltR\Delta dltA$  mutants (Table 3).  
335 Modification of *PdltR* to insert the CiaR binding site (*PdltR<sub>M1</sub>*) and of *PdltA* (*PdltA<sub>M1</sub>*) to  
336 insert the DltR binding site generated recombinant promoters that were no longer activated in  
337 a  $\Delta dltA$  genetic background, indicating that they were recognized neither by DltR nor by  
338 CiaR. These results support the hypothesis that these repeated motifs are essential for *PdltR*  
339 and *PdltA* promoter activity.

340

#### 341 **Cationic antimicrobial peptide sensitivity and acidic tolerance**

342 In *S. agalactiae*,  $\Delta dltA$  and  $\Delta ciaR$  mutants display an increased susceptibility to  
343 cationic antimicrobial peptides such as colistin, nisin, polymyxin B, and mammal defensins  
344 (Poyart et al., 2003, Quach et al., 2009). As *PdltA* is activated by CiaR, the susceptibility of a  
345  $\Delta ciaRH$  mutant might be due to a lower level of D-alanylation. To further investigate this  
346 hypothesis, the Etest method was used to accurately determine the minimal inhibitory  
347 concentration (MIC) of colistin against NEM316 WT and mutant derivatives (Fig. 4). The  
348 WT strain (MIC 32  $\mu\text{g/ml}$ ) and the  $\Delta relRS$  mutant (48  $\mu\text{g/ml}$ ) have similar MICs towards  
349 colistin but, as shown previously (Poyart et al., 2003), the  $\Delta dltA$  mutant is highly sensitive to  
350 this antibiotic (MIC 8  $\mu\text{g/ml}$ ). Interestingly, whereas the  $\Delta dltR$  and  $\Delta ciaRH$  mutants showed  
351 intermediate MIC values (16  $\mu\text{g/ml}$ ), the  $\Delta dltR\Delta ciaRH$  is as sensitive as the  $\Delta dltA$  mutant

352 (MIC 8  $\mu\text{g/ml}$ ). These results show that these two TCS act synergistically to activate the  
353 expression of the *dlt* operon of *S. agalactiae*.

354 For *Streptococcus mutans* it had been shown that inactivation of the *dlt* operon led to a  
355 loss of acid tolerance and that the mutant strain was unable to initiate growth below pH 6.5  
356 (Boyd et al., 2000). We therefore compared the acid sensitivity of NEM316 WT and its  
357 mutant derivatives after overnight growth in TH buffered at a neutral or acidic pH (7.3 and  
358 5.1, respectively). As shown in Fig. 5, all strains grow similarly at neutral pH whereas the  
359  $\Delta dltA$  and  $\Delta dltR\Delta ciaRH$  mutant strains have a severe growth defect at pH 5.3, the single  
360 mutants  $\Delta dltR$  and  $\Delta ciaRH$  show only a weak reduction of the final  $\text{OD}_{600}$  reached. This acid  
361 sensitivity parallels colistin sensitivity, indicating that it is related to the expression level of  
362 the *dltABCD* operon. We thus conclude that the level of D-alanylation of the LTAs of  
363 *S. agalactiae* is important both for resistance to cationic antimicrobial peptides and for growth  
364 in low pH conditions, two properties critical for carriage and invasion.

365

## 366 DISCUSSION

367

368 In most Firmicutes, D-alanylation of LTA is a critical modification of the cell  
369 envelope catalyzed by the proteins encoded in the *dltABCD* operon, which modulates the  
370 surface net charge. Modulation of the relative negative charge of the bacterial cell surface  
371 contributes not only to the sensitivity to cationic peptides, but also to the interaction with  
372 charged and non-charged surfaces, and to the recognition of the bacteria by the host defense  
373 mechanisms. It also interferes with the bacterial physiology as exemplified by our  
374 transcriptome analysis of the *S. agalactiae*  $\Delta dltA$  mutant that revealed a strong induction of  
375 the stress response. The level of D-alanylation of LTA has therefore to be tightly regulated.  
376 The transcriptional regulation of the *dlt* operon has been studied in diverse Firmicutes



377 revealing different regulatory strategies. In *Bacillus subtilis*, the *dlt* operon is under the  
378 control of the alternative sigma factor  $\sigma_D$  and is expressed in exponential phase (Perego et al.,  
379 1995). In pathogenic bacteria of the same phylogenetic lineage, the expression of the *dlt*  
380 operon is under the positive control of homologous TCSs: VirRS in *Listeria monocytogenes*  
381 (Mandin et al., 2005), GraRS in *Staphylococcus aureus* (Herbert et al., 2007), and ApsRS in  
382 *Staphylococcus epidermidis* (Li et al., 2007). In *S. aureus*, the situation is more complex as  
383 the *dlt* operon is also under the negative control of the AlsSR TCS (Koprivnjak et al., 2006)  
384 and was also identified as moderately up and downregulated in transcriptome profiling of *agr*  
385 and *rot* mutants, respectively (Dunman et al., 2001, Said-Salim et al., 2003). However,  
386 whether these regulations are direct or indirect is not yet known. In streptococci, the *dlt*  
387 operon has been shown to be under the control of different TCS, being down regulated in a  
388 *ciaR* pneumococcal mutant (Mascher et al., 2003) although the CiaR binding site was not  
389 identified (Halfmann et al., 2007). In *S. gordonii*, it is under the direct or indirect negative  
390 control of the LiaRS TCS (McCormick et al., 2011).

391 In *S. agalactiae*, the *dltABCD* operon has a peculiar organization as two genes  
392 encoding a TCS named *dltS* and *dltR* are located upstream the *dltA* gene (Poyart et al., 2001).  
393 Here, we showed that the expression of this locus is under the control of three promoters  
394 designated  $P_{0dltR}$ ,  $P_{dltR}$ , and  $P_{dltA}$ . The most upstream promoter,  $P_{0dltR}$  is constitutively  
395 expressed whereas the two others,  $P_{dltR}$  and  $P_{dltA}$  are strongly induced in the absence of D-  
396 alanylation revealing a feedback control loop for the homeostasis of the bacterial surface net  
397 charge. These two promoters are tightly regulated by two TCSs, DltRS and CiaRH, which are  
398 necessary for the induction in the absence of D-alanylation of  $P_{dltR}$  and  $P_{dltA}$ , respectively  
399 (Table 3). It is likely that these regulations are direct as we identified putative binding sites  
400 overlapping the  $-35$  promoter element for both regulators, a feature characteristic of  
401 promoters that are subject to Class II activation (Browning & Busby, 2004) (Fig. 1, 2 and 3).

402 Therefore, in *S. agalactiae* D-alanylation is under the dual control of two TCSs, DltRS and  
403 CiaRH, a complex genetic regulation unique among streptococci.

404 The diversity in the regulation of the *dlt* operon among streptococci and in particular  
405 its control by two TCSs in *S. agalactiae* raises the question of its evolutionary origin. Until  
406 recently, the *dltRS* genes were only found in *S. agalactiae* but we showed here that orthologs  
407 of these genes are present upstream the *dltABCD* operon of *S. infantarius*, a subspecies of the  
408 *bovis* group considered as distantly related to *S. agalactiae*. Alignment of the six *dlt* genes and  
409 Dlt proteins shows overall similar levels of conservation, which are higher than with their  
410 counterparts in *S. pyogenes*, a species considered as more closely related to *S. agalactiae*  
411 (Table 4). The genetic structure of the *dltRSABCD* loci in these two species is conserved (Fig.  
412 1), but analysis of the genomic context of the *dlt* loci in streptococci revealed no conservation  
413 of the surrounding genes between the different species, in particular between *S. agalactiae*  
414 and *S. infantarius* (Fig. 1). Therefore, this similar genetic organization likely results from an  
415 ancient acquisition of *dltRS* by the ancestor of one of the two species followed by the  
416 horizontal transfer of the whole locus and subsequent species divergence. The fact that  
417 *S. agalactiae* and *S. infantarius* are both regular inhabitants of the human gastrointestinal tract  
418 is consistent with this hypothesis. Comparison of the *PdltA* promoter regions among  
419 streptococci allows to hypothesize on the origin of the regulation by CiaRH in *S. agalactiae*.  
420 *PdltA* is located in the 3' part of the *dltS* gene in *S. agalactiae*, a region missing in other  
421 streptococci with the exception of *S. infantarius*. However, the alignment of the *PdltA*  
422 sequences in these two species revealed that, whereas the -10 promoter region is conserved,  
423 the predicted CiaR binding site is not found in *S. infantarius* (Fig. 2B). This suggests a more  
424 recent acquisition of this TCS regulation by *S. agalactiae* and a convergent evolution of  
425 *S. pneumoniae* and *S. agalactiae* for the regulation of D-alanylation by CiaRH. However, the  
426 CiaR regulation of *PdltA* might also have been lost in *S. infantarius*.

427           The absence of D-alanylation of LTA in *S. agalactiae* led to the sensitivity to cationic  
428 peptides and acidic pH (Fig. 4 and 5). Analysis of these two phenotypes in the *dltA* mutant  
429 strain and in the various regulatory systems studied here (DltRS, CiaRH, and RelRS) revealed  
430 a correlation between the expression level of the *dlt* operon and the sensitivity to these  
431 parameters. In particular, the  $\Delta$ *ciaRH* $\Delta$ *dltR* is as sensitive to cationic peptides and acidic pH  
432 as the *dltA* mutant. This observation indicates that the constitutive  $P_{0dltR}$  promoter might  
433 have only a minor contribution to the synthesis of the D-alanylation machinery. This  
434 constitutive promoter might direct the transcription of *dltRS* to ensure its expression in  
435 conditions where  $P_{dltR}$  is not activated by DltR. Its role would be to avoid the emergence of  
436 cells devoid of DltR that are unable to rapidly respond to D-alanyl ester deprivation of LTAs.  
437 Furthermore, our results demonstrating that  $P_{dltA}$  is regulated by CiaRH suggest that a  
438 reduced level of LTA D-alanylation may contribute to the reported sensitivity of the *ciaRH*  
439 mutant to cationic peptides and its reduced virulence in a mouse model (Quach et al., 2009).  
440 The combined sensitivity of the  $\Delta$ *dltRS* and  $\Delta$ *ciaRH* mutants to antimicrobial peptides and  
441 acidity emphasizes the important roles of these systems in GBS virulence, in particular for  
442 survival in macrophages and neutrophils.

443           An important issue in the study of the regulation by TCSs is the environmental signal  
444 perceived by the sensor protein. The inactivation of *dltA* led to the induction of three TCSs:  
445 DltRS, CiaRH, and RelRS, which in turn are responsible for 62 % of the transcriptional  
446 adaptation to the depletion of LTA D-alanylation (Table 1). Here we showed that the  
447 transcriptional activation of *dltRS* and *ciaRH* is probably a direct consequence of the absence  
448 of D-alanylation, whereas the up regulation of *relRS* is due to the activation of the CiaRH  
449 system. Interestingly, in *S. mutans*, CiaRH is activated in medium containing low calcium  
450 concentrations that are sensed by a small peptide encoded by *ciaX*, a gene located  
451 immediately upstream *ciaR* (He et al., 2008).  $Ca^{++}$  concentration may also alter the apparent

452 net charge of the surface by neutralizing negative charges of the cell envelop. The stimuli  
453 sensed by the CiaRH system in other streptococci, including GBS, which apparently do not  
454 code for a CiaX ortholog remain to be identified. Our working hypothesis is that DltS and  
455 CiaH sensors of *S. agalactiae* respond to different environmental changes related to surface  
456 charge modifications by controlling the level of D-alanylation of the LTAs. Taken together  
457 our results led us to propose a model in which D-alanylation of LTA in *S. agalactiae* is under  
458 the control of a dual feedback loop involving the two TCSs DltRS and CiaRH (Fig. 6). In this  
459 model, the level of D-alanylation modulates the activity of DltR and CiaR and the expression  
460 of the *dltABCD* operon is in turn regulated by these two regulators. This dual regulation  
461 ensures a fine-tuning of the expression of the *dlt* operon expression and consequently of the  
462 net-charge of the bacterial cell surface. Therefore in addition to its role in the modification of  
463 LTA, the *dlt* operon is central to the physiology of *S. agalactiae* controlled by CiaRH.

464 By deciphering the regulatory role of CiaR, we defined two classes of genes: those  
465 whose transcription is activated by CiaR only in absence of D-alanylation (e.g. *relPRS*) and  
466 those also activated in presence of D-alanylation (e.g. the ncRNAs). Comparative genomics  
467 showed that only CiaRH and the essential WalKR TCSs (Dubrac *et al.*, 2008) are ubiquitous  
468 among streptococci (data not shown). CiaRH contributes both to adaptation to acidic  
469 conditions and to host infection but comparison of available transcriptomic analysis revealed  
470 that its targets are highly diverse among streptococci. Among the 13 promoters controlled by  
471 CiaR in *S. agalactiae*, only those regulating the transcription of *ciaRH* (autoregulation), the  
472 small csRNAs, and *htrA* are conserved. HtrA, an extracellular protein with protease and  
473 chaperone activity, contributes to thermal, oxidative, and acid tolerance and is necessary for  
474 virulence of *S. pneumoniae* and *S. pyogenes* (Jones *et al.*, 2001, Sebert *et al.*, 2002, Kang *et*  
475 *al.*, 2010). In addition, alike in GBS, CiaR of *S. mutans* controls the expression of the *relPRS*  
476 operon (Wu *et al.*, 2010) and thereby the synthesis of the alarmone (p)ppGpp that mediates

477 the stringent stress response. Collectively, these data support the hypothesis that GBS CiaRH  
478 play a critical role in physiologic homeostasis and stress tolerance by linking (p)ppGpp  
479 metabolism, acid stress tolerance, and the response to D-alanyl ester deprivation.

480

481

## 482 **EXPERIMENTAL PROCEDURES**

483

484 **Bacterial strains and growth conditions.** *S. agalactiae* strains used in this work are listed in  
485 Supplementary Table 4. *E. coli* DH5 $\alpha$  (Invitrogen, Paisley, United Kingdom) was used for  
486 cloning experiments. *S. agalactiae* was grown in Todd-Hewitt (TH) broth or agar (Difco  
487 Laboratories, Detroit, MI) and *E. coli* in Luria-Bertani medium. *S. agalactiae* strains were  
488 grown in liquid culture at 37°C in 50 ml Falcon tubes without shaking. Antibiotics were used  
489 at the following concentrations: for *E. coli* - ampicillin, 100  $\mu$ g/mL; erythromycin,  
490 150  $\mu$ g/mL; for *S. agalactiae* - erythromycin, 10 $\mu$ g/mL.

491

492 **General DNA techniques.** Standard recombinant techniques were used for nucleic acid  
493 cloning and restriction analysis (Sambrook *et al.*, 1989). Plasmid DNA from *E. coli* was  
494 prepared by alkaline lysis using the Qiagen Miniprep plasmid kit. Genomic DNA of  
495 *S. agalactiae* was extracted using the Qiagen DNeasy kit (Qiagen, Hilden, Germany). The  
496 oligonucleotide primers used for PCR reactions were designed with Primer3 software  
497 (<http://primer3.sourceforge.net/>) and are listed in Supplementary Table 5. Sequencing  
498 reactions were performed using the Taq BigDye Terminator cycle sequencing kit and the  
499 3730 capillary sequencer (Applied Biosystems, Foster city, CA).

500

501 **Plasmids and mutant strain construction.** The primers used for plasmid constructions are  
502 listed in Supplementary Table 5. In-frame deletion mutants of the *ciaRH*, *relRS* genes were  
503 constructed by using splicing-by-overlap-extension PCR, as previously described (Dramsi *et*  
504 *al.*, 2006), and the oligonucleotide pairs *ciaR\_EcoR1-ciaR\_Int1* and *ciaH\_Int2-ciaH\_BamHI*,  
505 or 1398-*Eco*-1398-*int1* and 1397-*int2*-1397-*BamHI* respectively. Resulting *EcoRI-BamHI*  
506 amplicons were cloned into the thermosensitive shuttle plasmid pG<sup>+</sup>host5. Transformants and  
507 integrants were selected on antibiotic supplemented solid media as described above. For the  
508  $\Delta$ *ciaRH* and  $\Delta$ *relRS* in frame deletion, integrant strains were serially passaged in liquid  
509 medium at 30°C without erythromycin selection to allow the excision of the plasmid  
510 pG<sup>+</sup>host5, leaving in the chromosome either the  $\Delta$ *ciaRH*,  $\Delta$ *relRS* deletion or the wild type  
511 allele. Dilutions of the serially passaged cultures were tested by PCR and then for  
512 erythromycin sensitivity to identify clones in which pG<sup>+</sup>host5 was excised (Biswas *et al.*,  
513 1993). All in-frame deletions were confirmed by PCR and sequence analysis. The double  
514 mutant strain  $\Delta$ *ciaRH* $\Delta$ *dltA* and  $\Delta$ *relRS* $\Delta$ *dltA*, were obtained by mutating the *dltA* gene in the  
515  $\Delta$ *ciaRH* and  $\Delta$ *relRS* NEM316 derivative respectively as previously described (Poyart *et al.*,  
516 2001). The  $\Delta$ *dltR* $\Delta$ *ciaRH* strain was obtained by deleting the *ciaRH* genes in the  $\Delta$ *dltR* mutant  
517 strain.

518 *PdltA* activity was quantified by using the previously described plasmid  
519 *PdltA*pTCVlac expressing the corresponding *E. coli lacZ* fusion (Poyart *et al.*, 2001). For  
520 *PdltR* activity the *PdltR* fragment was amplified by using the oligonucleotide pair *PdltR-Eco* -  
521 *PdltR-Bam*. To test the function of the putative DltR and CiaR boxes, the *PdltR* and *PdltA*  
522 fragments were modified using two oligonucleotide pairs *PdltR-Eco* - *PdltR\_M0R* and  
523 *PdltR\_M0F* - *PdltR-Bam* for *PdltR*<sub>M0</sub>; *PdltR-Eco* - *PdltR\_M1R* and *PdltR\_M1F* - *PdltR-Bam*  
524 for *PdltR*<sub>M1</sub> and *PdltA-Eco* - *PdltA\_M1R* and *PdltA\_M1F* - *PdltA-Bam* for *PdltA*<sub>M1</sub>. To  
525 quantify the P<sub>0</sub>*dltR* activity the P<sub>0</sub>*dltR*+*PdltR*<sub>M0</sub> fragment was obtained using the two

526 oligonucleotide pairs P<sub>0</sub>*dltR*-Eco - *PdltR*\_M0R and *PdltR*\_M0F - *PdltR*-Bam. The resulting  
527 *EcoRI*-*Bam*H1 amplicons were cloned into the replicative shuttle vector pTCV-*lac* (Poyart &  
528 Trieu-Cuot, 1997). The recombinant plasmids were transformed into the WT NEM316 strain  
529 and the mutant derivatives by electroporation.

530

531 **β-galactosidase Assay.** *S. agalactiae* cells containing *lacZ* fusions were cultivated in TH  
532 medium at 37°C without shaking. Cells were collected in the mid-exponential phase (0.3-0.4  
533 OD), and β-galactosidase activity was assayed as described previously (Poyart et al., 2001).  
534 Each β-galactosidase activity assay was conducted in three replicates.

535

536 **Microarray experiments.** *S. agalactiae* mutant and wt strains were cultured in TH medium  
537 at 37C. Total RNA was extracted as previously described (Lamy *et al.*, 2004). *S. agalactiae*  
538 was harvested for RNA isolation at exponential growth phase (OD 0.35-0.4). RNA was  
539 prepared from three independent cultures and each RNA sample was hybridized twice to the  
540 microarrays (dye swap). RNA was reverse-transcribed with Superscript indirect cDNA kit  
541 (Invitrogen) and labeled with Cy5 or Cy3 (Amersham Biosciences) according to the supplier's  
542 instructions. The microarray contains 6835 60mer oligonucleotides specific for 2134  
543 predicted genes of the genome of NEM316 strain. The oligonucleotide design was carried out  
544 with the OligoArray server (Rouillard *et al.*, 2002) (<http://berry.engin.umich.edu/oligoarray/>)  
545 and with in house written scripts. The microarray was manufactured by Agilent Technologies.  
546 For hybridization, Cy3 and Cy5 target quantities were normalized. Arrays were scanned in an  
547 Axon 4200AL dual laser scanner connected to the GenePix Pro 6.0 image acquisition and  
548 analysis software package. Slides were scanned at 635nm and 532nm wavelength. An initial  
549 pre-scan of the slide was performed to determine success of hybridization and to allow  
550 defining the boundaries of the array features. PMT gain settings were adjusted at this step to

551 ensure that the overall red-to-green ratio was similar. Spots were excluded from the analysis  
552 in case of high local background fluorescence slide abnormalities or weak intensity. No  
553 background was subtracted. Data normalization and differential analyses were conducted  
554 using the R software (<http://www.r-project.org>). A global intensity-dependent normalization  
555 using the LOESS procedure (Yang *et al.*, 2002) was performed on a slide-by-slide basis  
556 (BioConductor package marray; <http://www.bioconductor.org/packages/bioc/html/marray.html>) to correct the dye bias. To determine differentially expressed genes, we  
557 performed a paired t test using the Benjamini and Yekutieli p-value correction method  
558 (Reiner *et al.*, 2003). For each gene 3 probes were present on the microarray. The cut off for  
559 the expression ratio was set to either superior/equal to 2 or inferior/equal to 0.5 and the  
560 general ratio of expression of each gene was calculated as the average expression ratio from  
561 the different probes. The complete dataset of our transcriptome analysis was submitted to the  
562 ArrayExpress database maintained at <http://www.ebi.ac.uk/microarray-as/ae/> under the  
563 Accession numbers: E-MEXP-3317, E-MEXP-3318, E-MEXP-3319. Putative CiaR and DltR  
564 binding motifs were identified by using the Search Pattern function of the SagaList web  
565 served (<http://genolist.pasteur.fr/SagaList/>).

567  
568 **RTqPCR.** For each strain mRNA was extracted in triplicates and reverse transcription was  
569 performed using Superscript II reverse transcriptase (Invitrogen) and random primers  
570 (Roche). Real-time PCR was performed on cDNA preparations using the SYBR green  
571 detection system (Applied Biosystems, Warrington, UK) and the ABI Prism 7900 sequence  
572 detector (Applied Biosystems) according to the manufacturer's instructions. Triplicate assays  
573 were performed for each point. The *S. agalactiae gyrA* transcripts were used as an internal  
574 control to normalize the amounts of RNA between the various samples. Standard curves to  
575 quantify *gyrA* mRNA and the csRNA under study were generated using the NEM316 wild-



576 type strain as a reference. To determine if the csRNA was differentially expressed a t test was  
577 performed using the R software, and a p-value < 0.05 was considered as significant. Primer  
578 sequences are listed in Table S5.

579

580 **Etest and Acid tolerance.** Etest (AB BIODISK) is a convenient assay to determine the  
581 minimal inhibitory concentration (MIC) of an antibiotic. Bacteria were inoculated in TH  
582 medium buffered at pH7 by Hepes 100mM. They were diluted 100 times after over-night  
583 growth at 37°C and 10 ml were poured onto TH agar plates. After 10 minutes, the excess of  
584 bacteria and medium was removed carefully and the plate was dried under a microbiology  
585 hood before applying the colistin Etest. The MIC was determined according to the  
586 recommendation of the manufacturer. Acidic tolerance was tested by comparing the OD<sub>600</sub>  
587 reached in TH medium at pH 5.1 and 7.3 after 18 hour cultures of the different strains  
588 inoculated at an OD<sub>600</sub> of 0.1.

589

## 590 **ACKNOWLEDGEMENTS**

591 This work was supported by the MNRT (ministère de la recherche et technologie) and  
592 Fondation pour la Recherche Médicale (V. D. C.) and by grant ANR-06-PATHO-001-01  
593 from the Agence Nationale de la Recherche "ERA-NET PathoGenoMics". We are most  
594 grateful to Carmen Buchrieser for critical reading of the manuscript and for fruitful  
595 discussions. We wish to thank Shaynoor Dramsi and Arnaud Firon for their constant interest  
596 and Corinne Genève for technical support.

597

## 598 **REFERENCES**

599 Abachin, E., Poyart, C., Pellegrini, E., Milohanic, E., Fiedler, F., Berche, P. & Trieu-Cuot, P., (2002)  
600 Formation of D-alanyl-lipoteichoic acid is required for adhesion and virulence of *Listeria*  
601 *monocytogenes*. *Mol. Microbiol.* **43**: 1-14.

- 602 Abi Khattar, Z., Rejasse, A., Destoumieux-Garzon, D., Escoubas, J. M., Sanchis, V., Lereclus, D., et  
603 al., (2009) The *dlt* operon of *Bacillus cereus* is required for resistance to cationic antimicrobial  
604 peptides and for virulence in insects. *J. Bacteriol.* **191**: 7063-7073.
- 605 Biswas, I., Gruss, A., Ehrlich, S. D. & Maguin, E., (1993) High-efficiency gene inactivation and  
606 replacement system for gram-positive bacteria. *J. Bacteriol.* **175**: 3628-3635.
- 607 Boyd, D. A., Cvitkovitch, D. G., Bleiweis, A. S., Kiriukhin, M. Y., Debabov, D. V., Neuhaus, F. C. &  
608 Hamilton, I. R., (2000) Defects in D-alanyl-lipoteichoic acid synthesis in *Streptococcus*  
609 *mutans* results in acid sensitivity. *J. Bacteriol.* **182**: 6055-6065.
- 610 Browning, D. F. & Busby, S. J., (2004) The regulation of bacterial transcription initiation. *Nat Rev*  
611 *Microbiol* **2**: 57-65.
- 612 Clemans, D. L., Kolenbrander, P. E., Debabov, D. V., Zhang, Q., Lunsford, R. D., Sakone, H., et al.,  
613 (1999) Insertional inactivation of genes responsible for the D-alanylation of lipoteichoic acid  
614 in *Streptococcus gordonii* DL1 (Challis) affects intrageneric coaggregations. *Infect. Immun.*  
615 **67**: 2464-2474.
- 616 Collins, L. V., Kristian, S. A., Weidenmaier, C., Faigle, M., Van Kessel, K. P., Van Strijp, J. A., et al.,  
617 (2002) *Staphylococcus aureus* strains lacking D-alanine modifications of teichoic acids are  
618 highly susceptible to human neutrophil killing and are virulence attenuated in mice. *J. Infect.*  
619 *Dis.* **186**: 214-219.
- 620 Cox, K. H., Ruiz-Bustos, E., Courtney, H. S., Dale, J. B., Pence, M. A., Nizet, V., et al., (2009)  
621 Inactivation of DltA modulates virulence factor expression in *Streptococcus pyogenes*. *PLoS*  
622 *ONE* **4**: e5366.
- 623 Debabov, D. V., Kiriukhin, M. Y. & Neuhaus, F. C., (2000) Biosynthesis of lipoteichoic acid in  
624 *Lactobacillus rhamnosus*: role of DltD in D-alanylation. *J. Bacteriol.* **182**: 2855-2864.
- 625 Dermer, P., Lee, C., Eggert, J. & Few, B., (2004) A history of neonatal group B streptococcus with its  
626 related morbidity and mortality rates in the United States. *J. Pediatr. Nurs.* **19**: 357-363.
- 627 Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prevost, M. C., Kojadinovic, M., et al., (2006)  
628 Assembly and role of pili in group B streptococci. *Mol. Microbiol.* **60**: 1401-1413.
- 629 Dubrac, S., Bisicchia, P., Devine, K. M. & Msadek, T., (2008) A matter of life and death: cell wall  
630 homeostasis and the WalKR (YycGF) essential signal transduction pathway. *Mol. Microbiol.*  
631 **70**: 1307-1322.
- 632 Dunman, P. M., Murphy, E., Haney, S., Palacios, D., Tucker-Kellogg, G., Wu, S., et al., (2001)  
633 Transcription profiling-based identification of *Staphylococcus aureus* genes regulated by the  
634 *agr* and/or *sarA* loci. *J. Bacteriol.* **183**: 7341-7353.
- 635 Ferretti, J. J., McShan, W. M., Ajdic, D., Savic, D. J., Savic, G., Lyon, K., et al., (2001) Complete  
636 genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc Natl Acad Sci U S A* **98**:  
637 4658-4663.
- 638 Halfmann, A., Kovacs, M., Hakenbeck, R. & Bruckner, R., (2007) Identification of the genes directly  
639 controlled by the response regulator CiaR in *Streptococcus pneumoniae*: five out of 15  
640 promoters drive expression of small non-coding RNAs. *Mol. Microbiol.* **66**: 110-126.
- 641 He, X., Wu, C., Yarbrough, D., Sim, L., Niu, G., Merritt, J., et al., (2008) The *cia* operon of  
642 *Streptococcus mutans* encodes a unique component required for calcium-mediated  
643 autoregulation. *Mol. Microbiol.* **70**: 112-126.
- 644 Herbert, S., Bera, A., Nerz, C., Kraus, D., Peschel, A., Goerke, C., et al., (2007) Molecular basis of  
645 resistance to muramidase and cationic antimicrobial peptide activity of lysozyme in  
646 staphylococci. *PLoS Pathog* **3**: e102.
- 647 Hyyrylainen, H. L., Vitikainen, M., Thwaite, J., Wu, H., Sarvas, M., Harwood, C. R., et al., (2000) D-  
648 Alanine substitution of teichoic acids as a modulator of protein folding and stability at the

- 649 cytoplasmic membrane/cell wall interface of *Bacillus subtilis*. *J. Biol. Chem.* **275**: 26696-  
650 26703.
- 651 Itou, H. & Tanaka, I., (2001) The OmpR-family of proteins: insight into the tertiary structure and  
652 functions of two-component regulator proteins. *J Biochem* **129**: 343-350.
- 653 Jayaraman, G. C., Penders, J. E. & Burne, R. A., (1997) Transcriptional analysis of the *Streptococcus*  
654 *mutans* *hrcA*, *grpE* and *dnaK* genes and regulation of expression in response to heat shock and  
655 environmental acidification. *Mol. Microbiol.* **25**: 329-341.
- 656 Jones, C. H., Bolken, T. C., Jones, K. F., Zeller, G. O. & Hruby, D. E., (2001) Conserved DegP  
657 protease in gram-positive bacteria is essential for thermal and oxidative tolerance and full  
658 virulence in *Streptococcus pyogenes*. *Infect. Immun.* **69**: 5538-5545.
- 659 Kajfasz, J. K., Rivera-Ramos, I., Abranches, J., Martinez, A. R., Rosalen, P. L., Derr, A. M., et al.,  
660 (2010) Two Spx proteins modulate stress tolerance, survival, and virulence in *Streptococcus*  
661 *mutans*. *J. Bacteriol.* **192**: 2546-2556.
- 662 Kang, K. H., Lee, J. S., Yoo, M. & Jin, I., (2010) The influence of HtrA expression on the growth of  
663 *Streptococcus mutans* during acid stress. *Mol Cells* **29**: 297-304.
- 664 Keefe, G. P., (1997) *Streptococcus agalactiae* mastitis: a review. *Can. Vet. J.* **38**: 429-437.
- 665 Koprivnjak, T., Mlakar, V., Swanson, L., Fournier, B., Peschel, A. & Weiss, J. P., (2006) Cation-  
666 induced transcriptional regulation of the *dlt* operon of *Staphylococcus aureus*. *J. Bacteriol.*  
667 **188**: 3622-3630.
- 668 Kovacs, M., Halfmann, A., Fedtke, I., Heintz, M., Peschel, A., Vollmer, W., et al., (2006) A functional  
669 *dlt* operon, encoding proteins required for incorporation of d-alanine in teichoic acids in gram-  
670 positive bacteria, confers resistance to cationic antimicrobial peptides in *Streptococcus*  
671 *pneumoniae*. *J. Bacteriol.* **188**: 5797-5805.
- 672 Kristian, S. A., Datta, V., Weidenmaier, C., Kansal, R., Fedtke, I., Peschel, A., et al., (2005) D-  
673 alanylation of teichoic acids promotes group A *streptococcus* antimicrobial peptide resistance,  
674 neutrophil survival, and epithelial cell invasion. *J. Bacteriol.* **187**: 6719-6725.
- 675 Lamy, M. C., Zouine, M., Fert, J., Vergassola, M., Couve, E., Pellegrini, E., et al., (2004) CovS/CovR  
676 of group B *streptococcus*: a two-component global regulatory system involved in virulence.  
677 *Mol. Microbiol.* **54**: 1250-1268.
- 678 Li, M., Lai, Y., Villaruz, A. E., Cha, D. J., Sturdevant, D. E. & Otto, M., (2007) Gram-positive three-  
679 component antimicrobial peptide-sensing system. *Proc Natl Acad Sci U S A* **104**: 9469-9474.
- 680 Mandin, P., Fsihi, H., Dussurget, O., Vergassola, M., Milohanic, E., Toledo-Arana, A., et al., (2005)  
681 VirR, a response regulator critical for *Listeria monocytogenes* virulence. *Mol. Microbiol.* **57**:  
682 1367-1380.
- 683 Marx, P., Nuhn, M., Kovacs, M., Hakenbeck, R. & Bruckner, R., (2010) Identification of genes for  
684 small non-coding RNAs that belong to the regulon of the two-component regulatory system  
685 CiaRH in *Streptococcus*. *BMC Genomics* **11**: 661.
- 686 Mascher, T., Zahner, D., Merai, M., Balmelle, N., de Saizieu, A. B. & Hakenbeck, R., (2003) The  
687 *Streptococcus pneumoniae* cia regulon: CiaR target sites and transcription profile analysis. *J.*  
688 *Bacteriol.* **185**: 60-70.
- 689 McCormick, N. E., Halperin, S. A. & Lee, S. F., (2011) Regulation of D-alanylation of Lipoteichoic  
690 Acid in *Streptococcus gordonii*. *Microbiology*.
- 691 Palumbo, E., Deghorain, M., Cocconcelli, P. S., Kleerebezem, M., Geyer, A., Hartung, T., et al.,  
692 (2006) D-alanyl ester depletion of teichoic acids in *Lactobacillus plantarum* results in a major  
693 modification of lipoteichoic acid composition and cell wall perforations at the septum  
694 mediated by the Acm2 autolysin. *J. Bacteriol.* **188**: 3709-3715.

- 695 Perego, M., Glaser, P., Minutello, A., Strauch, M. A., Leopold, K. & Fischer, W., (1995)  
696 Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in *Bacillus subtilis*.  
697 Identification of genes and regulation. *J. Biol. Chem.* **270**: 15598-15606.
- 698 Peschel, A., Otto, M., Jack, R. W., Kalbacher, H., Jung, G. & Gotz, F., (1999) Inactivation of the *dlt*  
699 operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other  
700 antimicrobial peptides. *J. Biol. Chem.* **274**: 8405-8410.
- 701 Phares, C. R., Lynfield, R., Farley, M. M., Mohle-Boetani, J., Harrison, L. H., Petit, S., et al., (2008)  
702 Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005.  
703 *JAMA* **299**: 2056-2065.
- 704 Poyart, C., Lamy, M. C., Boumaila, C., Fiedler, F. & Trieu-Cuot, P., (2001) Regulation of D-alanyl-  
705 lipoteichoic acid biosynthesis in *Streptococcus agalactiae* involves a novel two-component  
706 regulatory system. *J. Bacteriol.* **183**: 6324-6334.
- 707 Poyart, C., Pellegrini, E., Marceau, M., Baptista, M., Jaubert, F., Lamy, M. C. & Trieu-Cuot, P.,  
708 (2003) Attenuated virulence of *Streptococcus agalactiae* deficient in D-alanyl-lipoteichoic  
709 acid is due to an increased susceptibility to defensins and phagocytic cells. *Mol. Microbiol.*  
710 **49**: 1615-1625.
- 711 Poyart, C. & Trieu-Cuot, P., (1997) A broad-host-range mobilizable shuttle vector for the construction  
712 of transcriptional fusions to beta-galactosidase in gram-positive bacteria. *FEMS Microbiol.*  
713 *Lett.* **156**: 193-198.
- 714 Quach, D., van Sorge, N. M., Kristian, S. A., Bryan, J. D., Shelver, D. W. & Doran, K. S., (2009) The  
715 CiaR response regulator in group B *Streptococcus* promotes intracellular survival and  
716 resistance to innate immune defenses. *J. Bacteriol.* **191**: 2023-2032.
- 717 Reiner, A., Yekutieli, D. & Benjamini, Y., (2003) Identifying differentially expressed genes using  
718 false discovery rate controlling procedures. *Bioinformatics* **19**: 368-375.
- 719 Rouillard, J. M., Herbert, C. J. & Zuker, M., (2002) OligoArray: genome-scale oligonucleotide design  
720 for microarrays. *Bioinformatics* **18**: 486-487.
- 721 Said-Salim, B., Dunman, P. M., McAleese, F. M., Macapagal, D., Murphy, E., McNamara, P. J., et al.,  
722 (2003) Global regulation of *Staphylococcus aureus* genes by Rot. *J. Bacteriol.* **185**: 610-619.
- 723 Sambrook, J., Fritsch, E. F. & Maniatis, T., (1989) *Molecular Cloning, a laboratory manual, Second*  
724 *Edition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.
- 725 Seaton, K., Ahn, S. J., Sagstetter, A. M. & Burne, R. A., (2010) A Transcriptional Regulator and ABC  
726 Transporters Link Stress Tolerance, (p)ppGpp, and Genetic Competence in *Streptococcus*  
727 *mutans*. *J. Bacteriol.*
- 728 Sebert, M. E., Palmer, L. M., Rosenberg, M. & Weiser, J. N., (2002) Microarray-based identification  
729 of *htrA*, a *Streptococcus pneumoniae* gene that is regulated by the CiaRH two-component  
730 system and contributes to nasopharyngeal colonization. *Infect. Immun.* **70**: 4059-4067.
- 731 Silhavy, T. J., Kahne, D. & Walker, S., (2010) The bacterial cell envelope. *Cold Spring Harb Perspect*  
732 *Biol* **2**: a000414.
- 733 Wecke, J., Madela, K. & Fischer, W., (1997) The absence of D-alanine from lipoteichoic acid and wall  
734 teichoic acid alters surface charge, enhances autolysis and increases susceptibility to  
735 methicillin in *Bacillus subtilis*. *Microbiology* **143**: 2953-2960.
- 736 Weidenmaier, C., Kokai-Kun, J. F., Kristian, S. A., Chanturiya, T., Kalbacher, H., Gross, M., et al.,  
737 (2004) Role of teichoic acids in *Staphylococcus aureus* nasal colonization, a major risk factor  
738 in nosocomial infections. *Nat Med* **10**: 243-245.
- 739 Weidenmaier, C. & Peschel, A., (2008) Teichoic acids and related cell-wall glycopolymers in Gram-  
740 positive physiology and host interactions. *Nat Rev Microbiol* **6**: 276-287.

- 741 Wolz, C., Geiger, T. & Goerke, C., (2010) The synthesis and function of the alarmone (p)ppGpp in  
742 firmicutes. *Int J Med Microbiol* **300**: 142-147.
- 743 Wu, C., Ayala, E. A., Downey, J. S., Merritt, J., Goodman, S. D. & Qi, F., (2010) Regulation of  
744 *ciaXRRH* operon expression and identification of the CiaR regulon in *Streptococcus mutans*. *J.*  
745 *Bacteriol.* **192**: 4669-4679.
- 746 Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J. & Speed, T. P., (2002) Normalization  
747 for cDNA microarray data: a robust composite method addressing single and multiple slide  
748 systematic variation. *Nucleic Acids Res.* **30**: e15.

749  
750

751

## 752 FIGURE LEGENDS

753 **Fig. 1. Organization of the *S. agalactiae* *dlt* operon.** The organization of the *dlt* operon in  
754 *S. agalactiae* is compared to that in *S. infantarius* (Genbank NZ\_ABJK02000006), which contains  
755 orthologs of *dltR* and *dltS* and to that in *S. pyogenes* SF370 (Ferretti *et al.*, 2001) which does not.  
756 An hypothetical small gene, *dltX*, located between *dltS* and *dltA*, corresponds to the *dltS* - *dltA*  
757 intergenic region. *PdltR* and *PdltA*, previously described (Poyart *et al.*, 2001) and *P<sub>0</sub>dltR* identified  
758 in this work are indicated by arrows. Genes are indicated by arrows, *dltRS* genes are in light grey,  
759 *dltABCD* in dark grey and neighbouring genes, not orthologous in the three species in white.  
760 Putative transcription terminators are indicated by  $\bar{\cap}$ .

761

762 **Fig. 2. CiaRH dependent promoters.** A) DNA sequences of the promoters with predicted CiaR  
763 binding sites. Genes under the control of each promoter are indicated on the left, the repeats of the  
764 CiaR binding sites are in red with bases differing from the consensus (TTTAAG) in blue. -10  
765 promoter regions are in bold and underlined and the TSS in green. Numbers on the right panel are  
766 the ratios obtained in three transcriptome analyses: *DdltA*/wt, *DciaA*/wt and *DciaADdltA*/*DdltA*.  
767 "deleted" indicates that the *ciaA* gene is deleted in the tested strain. B) Alignment of the  
768 *S. agalactiae* *PdltA* region with the corresponding sequence in *S. infantarius*. Whereas the TSS  
769 and the -10 regions are conserved, the CiaR binding sites are not.

770

771 **Fig. 3. The *dltR* promoter region.** A) DNA sequence of the two promoters  $P_{0dltR}$  and  $P_{dltR}$   
772 identified upstream *dltR* in *S. agalactiae* and alignment with the corresponding region in  
773 *S. infantarius*. Only  $P_{dltR}$  is conserved. The repeats of the putative DltR binding sites are in blue  
774 with the base differing from the consensus (TTTCAA) in red. -10 promoter regions are in bold  
775 and underlined and the TSS in green. The -35 and -10 regions of  $P_{0dltR}$  are in orange. B)  
776 Alignment of *S. agalactiae*  $P_{dltR}$ ,  $P_{dltA}$  and the modified versions of these two promoters.  
777 Putative CiaR and DltR binding sites are in red and blue respectively. Repeats in  $P_{dltA_{M1}}$  have  
778 been modified to be identical to those in  $P_{dltR}$ , whereas repeats in  $P_{dltR_{M1}}$  were modified to be  
779 identical to those in  $P_{dltA}$ . The two modified nucleotides AA to TT in the second repeat of  
780  $P_{dltR_{M0}}$  are in orange.

781  
782 **Fig. 4: Resistance to colistin of NEM316 and mutant derivatives.** Colistin resistance was tested  
783 using the Etest. The strip contained a gradient of colistin concentrations allowing to determine the  
784 MIC of the different strains from the inhibitory zone. The genotypes of the tested strains are  
785 indicated above the pictures. The experiment shown is representative of three independent  
786 experiments. The MIC value corresponds to the value directly determined by the Etest.

787  
788 **Fig. 5: Effect of the pH on growth of strain NEM316 and mutant derivatives.** The effect of  
789 growth at pH 7.3 or 5.1 was determined as the OD600 reached after 18 hours of incubation at  
790 37 °C in TH medium buffered at the two pH.

791  
792 **Fig. 6: Regulatory model of LTA D-alanylation homeostasis.** In this model LTA D-alanylation  
793 lowers the activation of DltR and CiaR, as indicated by minus signs. These two regulators activate  
794 their own expression and the expression of the *dltA-D* genes as indicated by the plus signs. In

795 addition CiaR activates the expression of the *relPRS* operon. Genes are indicated by boxed arrows  
796 and regulatory genes are depicted in light grey.  
797

For Peer Review

797 Table 1: Genes differentially expressed in the  $\Delta dltA$  mutant compared to the WT strain

Gene	$\Delta dltA/wt$	$\Delta dltA\Delta dltR/\Delta dltA$	$\Delta dltA\Delta ciaRH/\Delta dltA$	$\Delta dltA\Delta reIS/\Delta dltA$	Annotation
Genes potentially regulated by DltR in response to the absence of D-alanylation					
<sup>a</sup> <i>gbs0213</i>	2.70	0.41	-	-	conserved protein of unknown function
<i>gbs0214</i>	3.07	0.36	-	-	conserved protein of unknown function
<i>gbs0215</i>	2.63	0.45	-	-	replication initiation factor, RepA family
<i>gbs0569</i>	2.72	0.48	-	-	acetoin reductase
<i>gbs0574</i>	0.45	2.76	-	-	inosine-uridine preferring nucleoside hydrolase family protein
<i>fbsA</i>	2.31	0.41	-	-	fibrinogen-binding protein, putative
<i>gbs1286</i>	6.18	0.31	-	-	protein of unknown function
<i>clpP</i>	2.57	0.49	-	-	ATP-dependent Clp protease proteolytic subunit
<i>gbs1805</i>	2.35	0.42	-	-	CHAP domain protein
<i>dltA</i>	2.87	0.35	-	-	D-alanine--D-alanyl carrier protein ligase
<i>dltS</i>	2.88	0.22	-	-	sensor histidine kinase DltS
<i>dltR</i>	2.47	<sup>b</sup> 0.15	-	-	DNA-binding response regulator DltR
<i>gbs1874</i>	2.02	0.35	-	-	alkyl hydroperoxide reductase, subunit C
<i>gbs1950</i>	2.90	0.40	-	-	phosphate ABC transporter, ATP-binding protein
<i>gbs1951</i>	2.79	0.34	-	-	phosphate ABC transporter, permease protein PtsA
<i>gbs1952</i>	3.28	0.35	-	-	phosphate ABC transporter, permease protein PstC
<i>gbs1953</i>	3.01	0.34	-	-	phosphate ABC transporter, phosphate-binding protein
<i>gbs2106</i>	2.78	0.42	-	-	transglycosylase-like domain protein
Genes potentially regulated by CiaR in response to the absence of D-alanylation					
<i>gbs0030</i>	1.88	-	0.12	-	similar to <i>Streptococcus equi</i> zocin A endopeptidase
<i>gbs0196</i>	4.55	-	0.01	-	oxidoreductase, NAD-binding
<i>ciaH</i>	2.19	-	<sup>c</sup> Deleted	-	sensor histidine kinase CiaH
<i>ciaR</i>	2.24	-	Deleted	-	DNA-binding response regulator CiaR
<i>reIS</i>	2.16	-	0.42	Deleted	sensor histidine kinase
<i>reIR</i>	2.52	-	0.44	Deleted	DNA-binding response regulator
<i>reIP</i>	2.96	-	0.31	-	RelA/SpoT domain protein
<i>gbs1415</i>	3.03	-	0.32	-	Predicted membrane protein
<i>gbs1725</i>	3.67	-	0.27	-	oxidoreductase, aldo/keto reductase family
<i>gbs1995</i>	5.92	-	0.09	-	PAP2 family protein
<i>gbs2003</i>	2.65	-	0.02	-	conserved protein of unknown function

CiaRH regulon



<i>htrA</i>	4.39	-	0.11	-	serine peptidase HtrA
<i>gbs0221</i>	3.59	-	2.35	-	RelE/ParE family protein
<i>gbs0472</i>	2.19	-	2.23	-	conserved protein of unknown function
<i>gbs1192</i>	2.04	-	2.00	-	aldehyde dehydrogenase family protein

Genes potentially regulated by RelR in response to the absence of D-alanylation

<i>gbs2083</i>	5.20	-	-	2.48	putative arginine transporter
<i>gbs2084</i>	10.21	-	-	3.49	carbamate kinase
<i>gbs2085</i>	7.68	-	0.49	3.47	ornithine carbamoyltransferase

Genes potentially regulated by RelRS and CiaRH

<i>rpsT</i>	4.17	-	2.45	2.79	30S ribosomal subunit protein S20
<i>gbs1778</i>	2.32	-	3.64	3.69	universal stress family protein

Genes potentially regulated by RelRS and DltR

<i>gbs0661</i>	2.52	0.31	-	2.46	putative DNA-entry nuclease
<i>gbs1116</i>	0.48	2.03	-	4.40	xanthine permease
<i>gbs1734</i>	2.45	0.40	-	0.5	PTS system, sucrose-specific IIABC component

Genes potentially regulated by DltR and CiaRH

<i>gbs0222</i>	4.12	0.45	2.59	-	protein of unknown function
<i>gbs0326</i>	3.47	0.40	2.55	-	ribosomal subunit interface protein
<i>gbs0600</i>	6.13	0.49	2.49	-	conserved protein of unknown function
<i>gbs0852</i>	3.40	0.46	2.53	-	protein of unknown function
<i>clpL</i>	6.48	0.28	4.58	-	Heat shock protein ClpL
<i>gbs1772</i>	2.11	0.46	2.76	-	formate acetyltransferase 1
<i>gbs1999</i>	3.16	0.43	6.19	-	rhodanese-like domain protein
<i>cfa</i>	3.37	0.35	5.21	-	CAMP factor
<i>gbs1202</i>	2.35	0.49	2.36	-	putative alkaline shock protein 23
<i>gbs1204</i>	2.04	0.45	2.35	-	putative alkaline shock protein 23
<i>gbs1695</i>	2.00	0.45	4.23	-	dihydroxyacetone kinase DAK2 domain protein
<i>gbs1697</i>	2.25	0.40	3.08	-	glycerol uptake facilitator protein
<i>hrcA</i>	2.72	0.30	2.32	-	heat-inducible transcription repressor
<i>grpE</i>	3.08	0.30	3.62	-	co-chaperone protein GrpE
<i>dnaK</i>	2.57	0.25	4.10	-	molecular chaperone DnaK
<i>dnaJ</i>	2.05	0.33	2.68	-	co-chaperone protein DnaJ
<i>clpE</i>	2.84	0.34	2.96	-	ATP-dependent Clp protease, ATP-binding subunit ClpE
<i>groEL</i>	2.25	0.32	2.63	-	chaperonin GroEL
<i>groES</i>	2.54	0.31	2.46	-	chaperonin GroES
<i>gbs1683</i>	0.44	2.00	0.37	-	branched chain amino acid:H <sup>+</sup> symporter
<i>gbs0189</i>	0.01	54.80	75.44	-	PTS system, trehalose-specific II ABC component

HrcA regulon

<i>gbs0190</i>	0.03	30.21	46.90	-	alpha amylase family protein
----------------	------	-------	-------	---	------------------------------

Gene potentially regulated by the three TCSs

<i>gbs2107</i>	4.91	0.32	0.46	0.46	LysM domain protein
----------------	------	------	------	------	---------------------

Genes that are not regulated by one of the three TCSs

<i>rpmV</i>	2.31	-	-	-	50S ribosomal protein L28
<i>gbs0158</i>	3.14	-	-	-	protein of unknown function
<i>gbs0168</i>	2.31	-	-	-	conserved protein of unknown function
<i>gbs0182</i>	2.35	-	-	-	LrgA family protein
<i>rpsO</i>	2.32	-	-	-	30S ribosomal subunit protein S15
<i>gbs0218</i>	2.32	-	-	-	protein of unknown function
<i>gbs0223</i>	2.20	-	-	-	conserved protein of unknown function
<i>gbs0257</i>	2.36	-	-	-	isoprenylcysteine carboxyl methyltransferase family protein
<i>gbs0358</i>	2.81	-	-	-	exported protein of unknown function
<i>gbs0397</i>	2.51	-	-	-	exported protein of unknown function
<i>gbs0471</i>	2.24	-	-	-	conserved protein of unknown function
<i>gbs0473</i>	2.25	-	-	-	exported protein of unknown function
<i>gbs0480</i>	2.02	-	-	-	protein of unknown function
<i>gbs0533</i>	2.56	-	-	-	protein of unknown function
<i>gbs0536</i>	4.46	-	-	-	conserved protein of unknown function
<i>gbs0551</i>	6.61	-	-	-	DNA-binding protein HU
<i>gbs0552</i>	2.45	-	-	-	protein of unknown function
<i>gbs0581</i>	2.05	-	-	-	putative ribosomal protein
<i>cylX</i>	2.14	-	-	-	<i>cylX</i> protein
<i>cylZ</i>	2.00	-	-	-	<i>cylZ</i> protein
<i>gbs0687</i>	2.49	-	-	-	conserved protein of unknown function
<i>gbs0761</i>	3.09	-	-	-	protein of unknown function
<i>gbs0764</i>	4.41	-	-	-	conserved protein of unknown function
<i>gbs0793</i>	2.03	-	-	-	protein of unknown function
<i>sod</i>	2.02	-	-	-	superoxide dismutase
<i>gbs0819</i>	2.07	-	-	-	glucosamine-6-phosphate isomerase
<i>prsA</i>	2.04	-	-	-	foldase protein PrsA
<i>gbs0838</i>	2.08	-	-	-	glutaredoxin-like protein
<i>gbs0932</i>	2.00	-	-	-	signal peptidase I
<i>gbs0951</i>	2.12	-	-	-	conserved exported protein of unknown function
<i>gbs1074</i>	2.29	-	-	-	conserved protein of unknown function
<i>gbs1085</i>	2.24	-	-	-	ATP cone domain protein
<i>gbs1111</i>	3.90	-	-	-	4-oxalocrotonate tautomerase
<i>gbs1210</i>	2.53	-	-	-	thioesterase family protein

<i>gbs1226</i>	2.43	-	-	-	protein of unknown function
<i>gbs1400</i>	2.13	-	-	-	efflux ABC transporter, permease/ATP-binding protein
<i>gbs1402</i>	2.25	-	-	-	transcriptional regulator, MarR family
<i>gbs1446</i>	2.33	-	-	-	protein of unknown function
<i>gbs1457</i>	2.62	-	-	-	putative ferredoxin
<i>rpsU</i>	5.25	-	-	-	30S ribosomal protein S21
<i>gbs1505</i>	3.87	-	-	-	protein of unknown function
<i>gbs1632</i>	0.50	-	-	-	leucine transporter ABC superfamily
<i>gbs1633</i>	4.49	-	-	-	protein of unknown function
<i>gbs1653</i>	2.82	-	-	-	conserved membrane protein of unknown function
<i>gbs1654</i>	2.35	-	-	-	protein of unknown function
<i>gbs1721</i>	2.16	-	-	-	universal stress family protein
<i>gbs1766</i>	2.19	-	-	-	protein of unknown function
<i>gbs1791</i>	2.78	-	-	-	conserved protein of unknown function
<i>rpmH</i>	3.17	-	-	-	50S ribosomal subunit protein L34
<i>gbs1904</i>	2.64	-	-	-	preprotein translocase subunit YajC
<i>gbs1972</i>	2.39	-	-	-	putative transcriptional regulator
<i>cspA</i>	0.48	-	-	-	cell surface serine endopeptidase CspA
<i>spxA</i>	2.17	-	-	-	transcriptional regulator SpxA
<i>gbs2053</i>	3.52	-	-	-	stress protein, member of the CspA-family
<i>gbs2055</i>	2.04	-	-	-	arginine repressor
<i>gbs2057</i>	2.13	-	-	-	protein of unknown function

798 <sup>a</sup> Except for *gbs0030*, only ratio lower than 0.5 or higher than 2 are given.

799 <sup>b</sup> The signal corresponds to the two probes matching sequences left in the  $\Delta dltR$  gene.

800 <sup>c</sup> deleted, means that the gene is deleted in the tested strains, no signal was detected.

801

801 Table 2. <sup>a</sup>Expression of csRNAs by RT-qPCR

	<b>csRNA10</b>	<b>csRNA11</b>
<b>NEM316</b>	0.91 (0.08)	0.72 (0.15)
<b><math>\Delta</math><i>ciaRH</i></b>	0.38 (0.08) <sup>b</sup>	0.28 (0.13) <sup>b</sup>
<b><math>\Delta</math><i>dltA</i></b>	2.47 (0.27) <sup>b</sup>	2.52 (0.71) <sup>b</sup>
<b><math>\Delta</math><i>CiaRH</i><math>\Delta</math><i>dltA</i></b>	0.11 (0.03) <sup>c</sup>	0.08 (0.05) <sup>c</sup>

802 <sup>a</sup> Values correspond to the relative mRNA concentration with the *gyrA* gene as control.

803 Standard deviation are indicated in parentheses.

804 <sup>b</sup> The ratio is significantly different to that observed in the wild-type.805 <sup>c</sup> The ratio is significantly different to that observed in the  $\Delta$ *dltA* mutant.

806 Table 3 Promoter activity of P<sub>0</sub>*dltR*, P*dltR*, and P*dltA* in *S. agalactiae* NEM316 and mutant derivatives

Promoter fused to <i>spoVG-lacZ</i> <sup>a</sup>	β-galactosidase activity (mean arbitrary units ± SD) <sup>b</sup>							
	NEM316	<i>ΔdltR</i>	<i>ΔCiaRH</i>	<i>ΔrelRS</i>	<i>ΔdltA</i>	<i>ΔdltRΔdltA</i>	<i>ΔCiaRHΔdltA</i>	<i>ΔrelRSΔdltA</i>
P <i>dltR</i>	33.5 (7.2)	1.45 (0.5)	27.8 (9.5)	24.1 (5.8)	302.9 (88.3)	4.4 (2.1)	279.4 (54.4)	248.4 (22.3)
P <i>dltR</i> <sub>M0</sub>	2.5 (1.7)	-	-	-	1.82 (1.8)	3.1 (2.5)	3.5 (2.3)	-
P <i>dltR</i> <sub>M1</sub>	2.4 (0.6)	1.2 (0.8)	-	2.4 (0.5)	3.3 (0.7)	1.1 (1.9)	2.3 (1.4)	2.3 (1.3)
P <sub>0</sub> <i>dltR</i> -P <i>dltR</i> <sub>M0</sub>	31.5 (7.2)	-	-	-	26.4 (8.3)	45.4 (6.2)	35.4 (2.6)	-
P <i>dltA</i>	4.0 (1.8)	2.2 (0.4)	ND	1.91 (1.7)	31.3 (4.2)	18.4 (9.5)	ND	16.6 (0.8)
P <i>dltA</i> <sub>M1</sub>	1.8 (1.2)	-	-	-	2.2 (1.2)	1.5 (2.2)	1.2 (0.4)	-

807

808 <sup>a</sup> DNA fragments containing the promoters (P*dltR*, P*dltR*<sub>M0</sub>, P*dltR*<sub>M1</sub>, P<sub>0</sub>*dltR*-P*dltR*<sub>M0</sub>, P*dltA*, and P*dltA*<sub>M1</sub>) were amplified by PCR and cloned into  
809 pTCV-*lac*.

810 <sup>b</sup> β-Galactosidase was assayed in cells collected in mid-exponential phase. Experiments were performed in triplicates and the mean values ± one  
811 standard deviation (SD) are indicated. ND, no detectable activity; -, not tested.

812 **Table 4:** Sequence comparison of *S. agalactiae dlt* genes with *S. infantarius* and *S. pyogenes*

	<sup>a,b</sup> <i>ciaR</i>	<i>dltR</i>	<i>dltS</i>	<i>dltA</i>	<i>dltB</i>	<i>dltC</i>	<i>dltD</i>
<i>S. infantarius</i>							
amino acid	87 %	72 %	56 %	82 %	76 %	87 %	61 %
nucleotide	77 %	64 %	62 %	74 %	71 %	82 %	65 %
<i>S. pyogenes</i> SF370							
amino acid	80 %	-	-	73 %	57 %	72 %	50 %
nucleotide	77 %			68 %	67 %	68 %	58 %

813 <sup>a</sup> Percentage of identities by BLASTP and BLASTN at the amino acid and nucleotide levels

814 respectively.

815 <sup>b</sup> Alignments with *ciaR* was used as a control

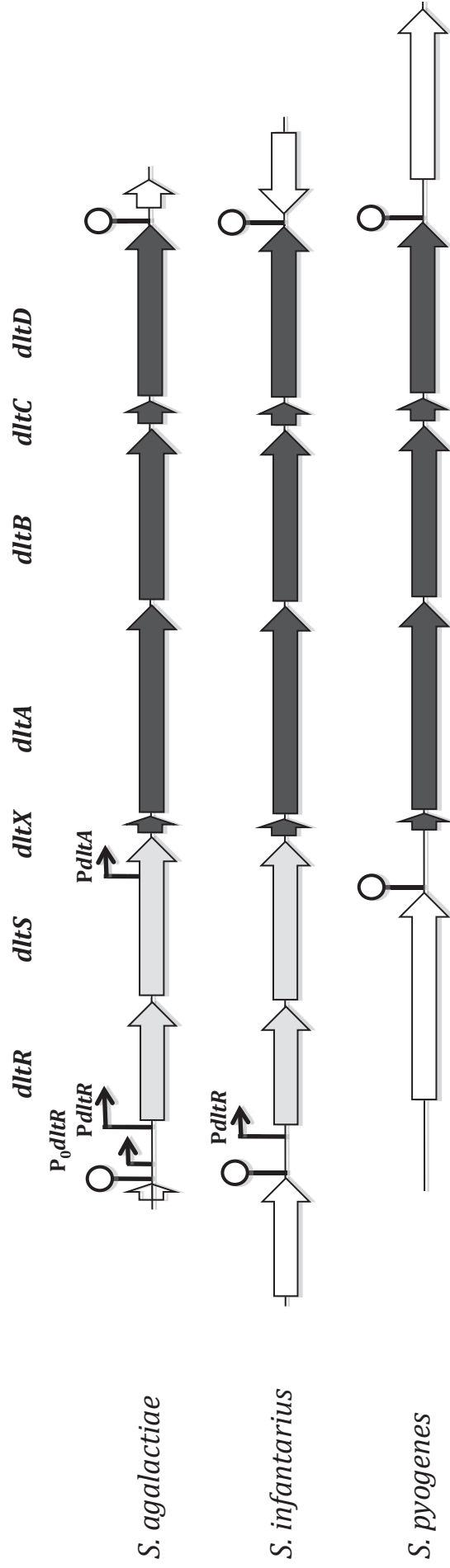


Figure 1







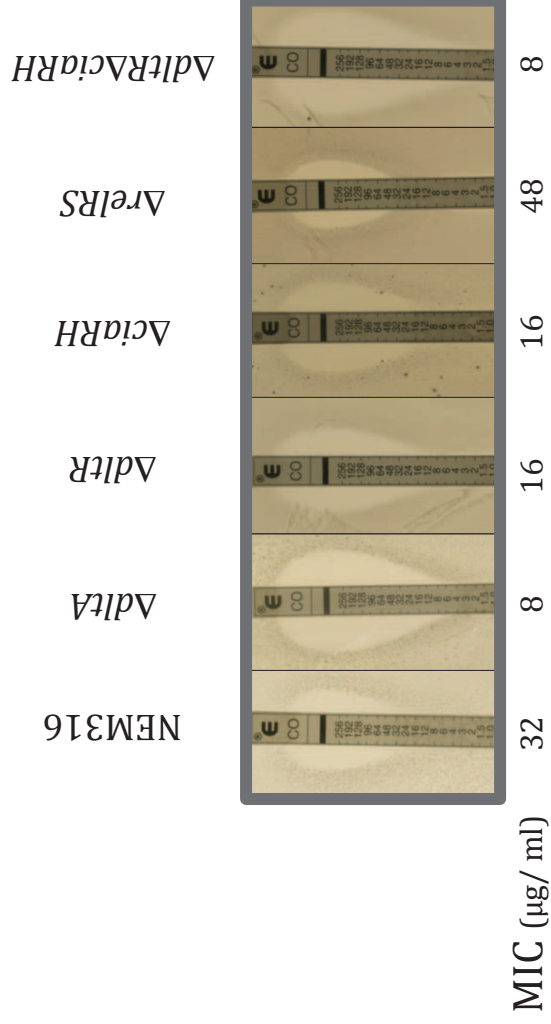


Figure 4

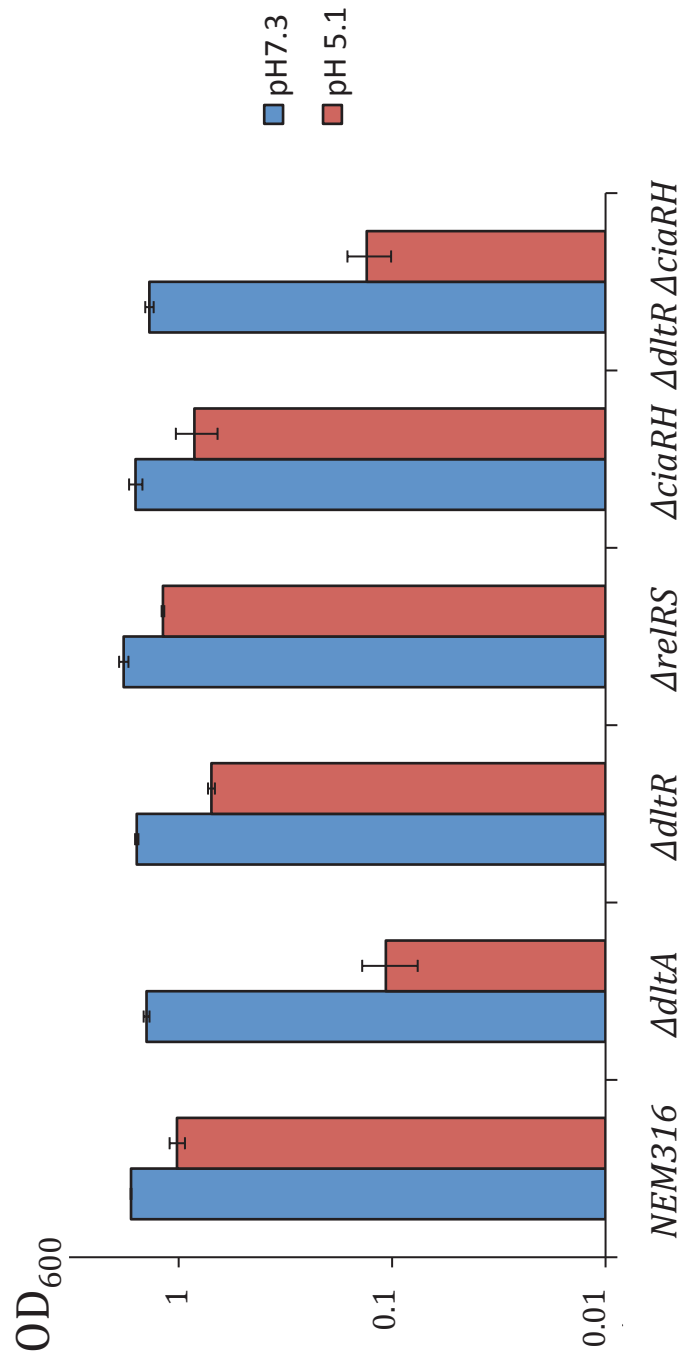


Figure 5

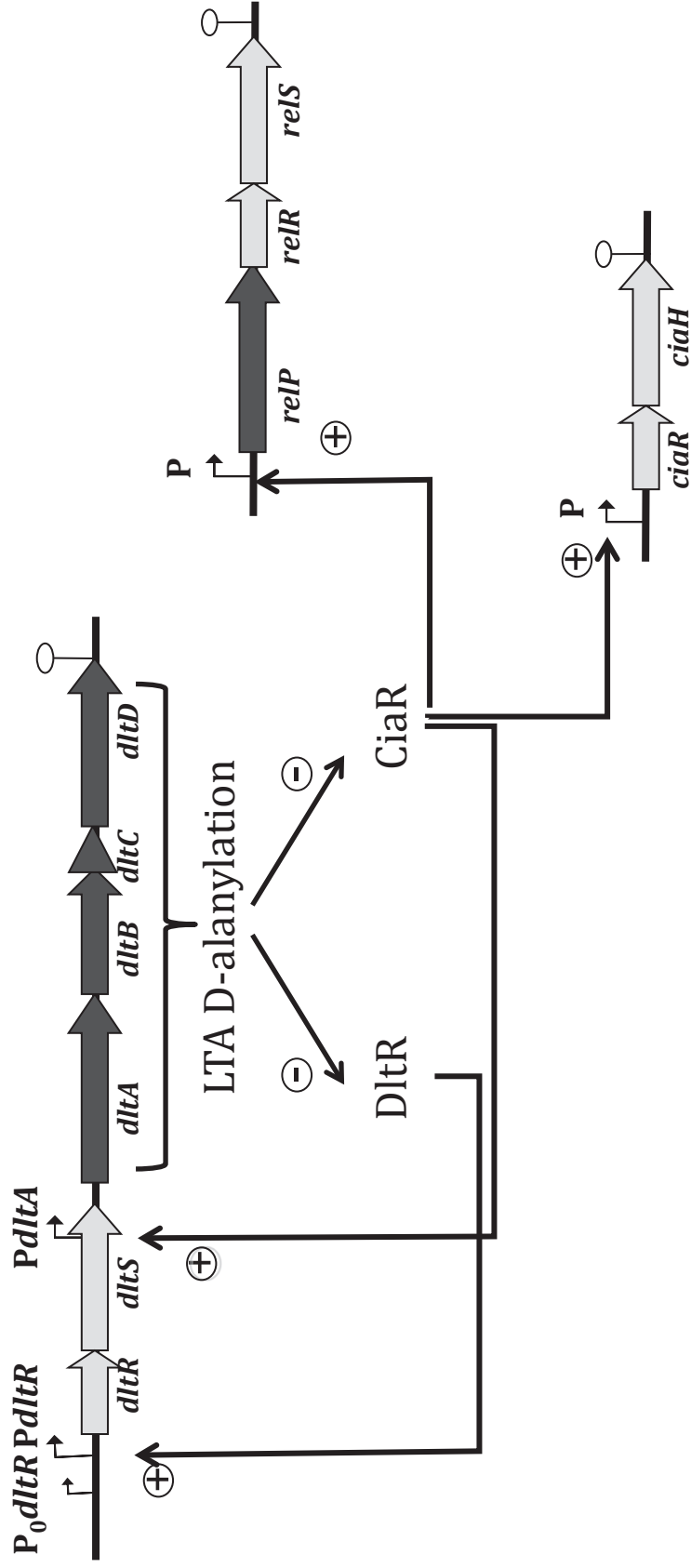
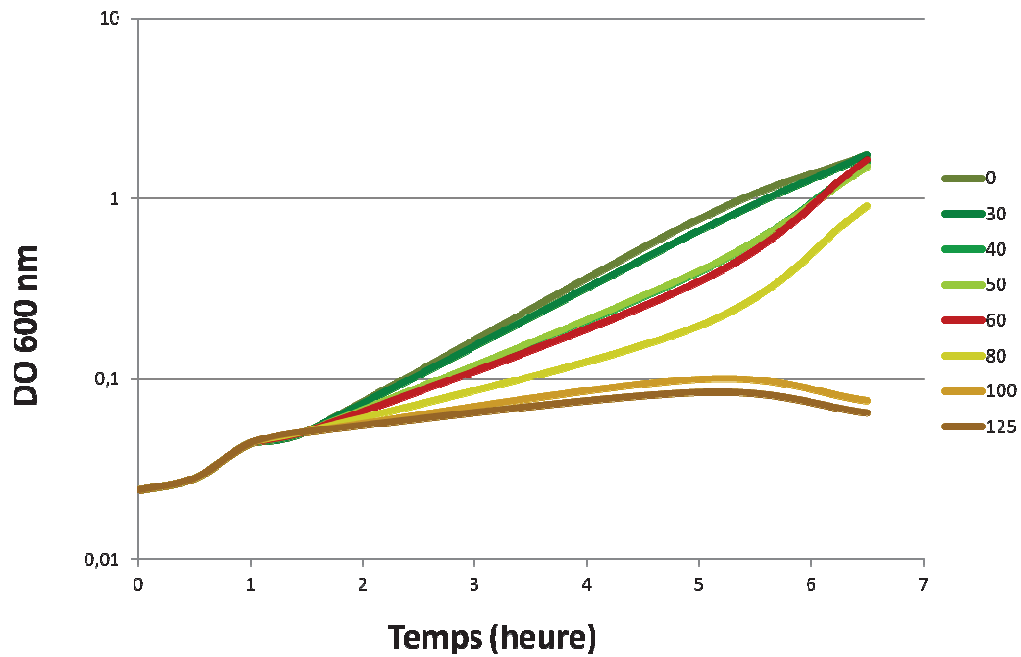


Figure 6

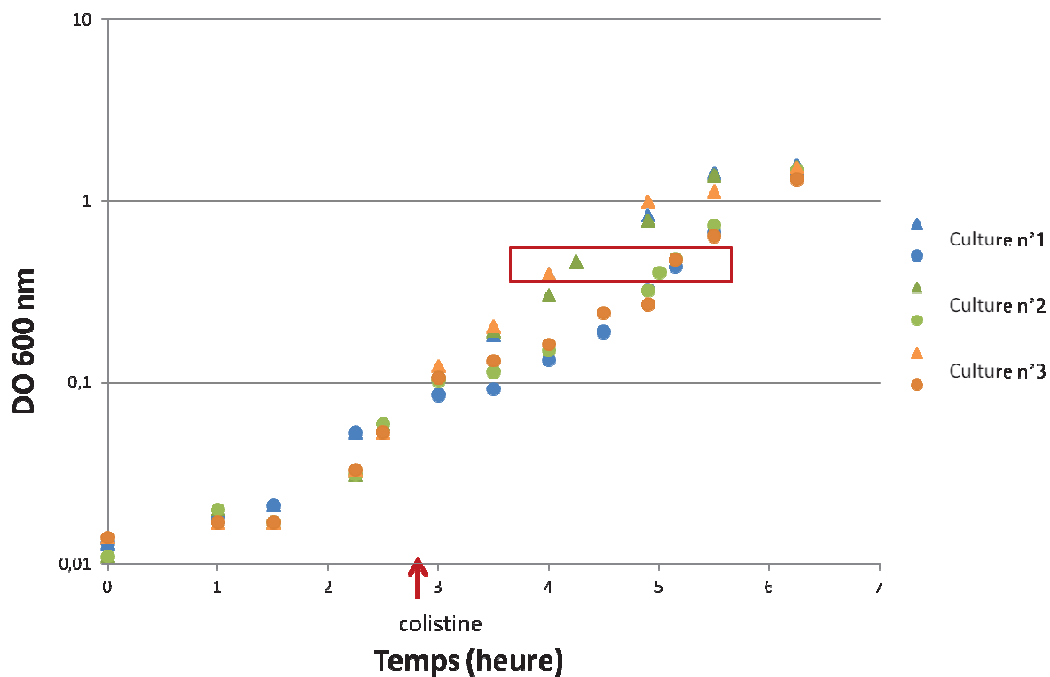
## **(2) Résultats complémentaires**

### **(a) Effets de concentrations sub-inhibitrices de colistine**

La colistine est un antibiotique polypeptidique cyclique, de la famille des polymyxines, produit par *Bacillus polymyxa*, et chargé positivement, mais ressemblant à un peptide cationique. Il agit comme détergent sur la membrane cellulaire des bactéries [335]. Afin d'étudier la réponse de *S. agalactiae* à la présence de peptides antimicrobiens, j'ai réalisé une analyse transcriptomique dans des conditions où la concentration de colistine est sub-inhibitrice afin de limiter la mise en place d'une réponse majoritairement liée à un fort ralentissement de la croissance. J'ai, dans un premier temps, déterminé la concentration sub-inhibitrice en analysant les effets sur la croissance d'une gamme de concentrations de colistine (Figure 16). Puis dans un second temps, j'ai étudié l'effet sur le profil d'expression d'une concentration de 60 µg/mL de colistine n'induisant qu'un faible ralentissement de la croissance (Figure 16). La colistine a été ajoutée à la culture à une  $DO_{600}$  à peu près égale à 0,05 après environ deux doublements, les bactéries ont été prélevées à une  $DO_{600}$  de 0,4 correspondant à environ trois doublements après l'ajout de l'antibiotique. Dans ces conditions on peut considérer que la réponse analysée correspond à une adaptation après ajout de colistine et non à la réponse au stress induit. Cette analyse a été faite en triplicats.



**Figure 16: Effet des différentes concentrations de colistine sur la croissance de la souche NEM316**  
 Sur ce graphique ne sont représentées que les courbes de croissance de la souche NEM316 en absence de colistine et en présence des concentrations finales de colistine de 30, 40, 50, 60, 80, 100 et 125 µg/mL. La courbe de croissance de NEM316 en présence de 60 µg/mL de colistine est indiquée en orange.



**Figure 17 Effets de doses sub-inhibitrices de colistine sur la croissance de la souche NEM316.**  
 Courbe de croissance en absence et après addition de colistine (60 µg/mL). Les cultures réalisées en absence ou en présence de colistine sont respectivement indiquées par des triangles et des ronds. Les couleurs correspondent à trois cultures indépendantes.

En réponse à la présence de colistine dans le milieu, 219 gènes sont différentiellement exprimés (Table 5), soit environ 10% des gènes de la souche NEM316. Parmi les 111 gènes réprimés, 24 codent pour des protéines impliquées dans le métabolisme énergétique (*atpB*, *atpF*, *atpH*), la machinerie de réplication- réparation- recombinaison (*rpoA*, *polA*, *recN*, *nrdE*, *deoD2*), le ribosome (14 gènes codant pour des protéines ribosomales) et la division cellulaire (*divIB*, *ftsZ*). La diminution de la transcription de ces gènes est probablement liée à la réduction de la vitesse de croissance de *S. agalactiae* en présence de colistine (Figure 17). L'expression de 6 régulateurs transcriptionnels est modifiée. La majorité de ces régulateurs est induite: Gbs0048 (famille CroCI), Gbs0685 (famille OmpR), Gbs1582 (famille LysR), Gbs1849 (famille LacI), RgfA; seul Gbs0546 (famille ArgR) est réprimé.

De façon surprenante aucune expression différentielle n'est observée, aussi bien pour l'opéron *dlt* que pour les régulateurs *ciaR* et *dltR*. Il n'y a pas non plus d'induction des protéines de stress induites dans un contexte  $\Delta dltA$ . La comparaison de ce transcriptome avec celui du mutant  $\Delta dltA$ /sauvage, indique que 28 des gènes différentiellement exprimés en réponse à la colistine le sont également dans le mutant  $\Delta dltA$ . La majorité de ces gènes est induite dans les deux transcriptomes. Parmi ces gènes trois correspondent à des gènes dont la transcription est activée par CiaR (*gbs0196*, *gbs1415*, *gbs1995*). L'induction de plusieurs cibles directes de CiaRH en présence de doses sub-inhibitrices suggère que le régulateur CiaR serait activé, mais que cette activation ne s'accompagnerait pas d'une augmentation de la transcription de *ciaR*. Cette activation serait partielle et n'entraînerait la transcription que des promoteurs les plus sensibles. Dans ce contexte il serait intéressant de déterminer si cette augmentation d'activité passe par une modification de l'état de phosphorylation du régulateur CiaR.

Le fait marquant de cette analyse est la répression de fonctions liées au taux de croissances en l'absence d'induction des protéines de stress. Une hypothèse pourrait être que la colistine ralentisse l'import de nutriments, ou affecte le niveau énergétique de la cellule. L'analyse des autres fonctions modifiées ne permet pas d'identifier un mécanisme spécifique d'adaptation à l'ajout de colistine en conditions sub-inhibitrices.

Chez *S. pneumoniae*, une analyse portant sur la réponse transcriptomique à différents peptides antimicrobiens (nisine, LL-37, et bacitracine) a permis de montrer que la résistance de *S. pneumoniae* à ces peptides fait intervenir des ABC transporteurs [336]. Or dans les données obtenues chez *S. agalactiae*, l'activation de l'expression d'un unique ABC transporteur est

observée. Cet ABC transporteur n'est orthologue à aucun des transporteurs identifiés dans l'étude réalisée chez *S. pneumoniae*. Cette observation suggère la mise en place d'un autre mécanisme de résistance à de faibles concentrations de colistine chez *S. agalactiae*.

La colistine est un antibiotique qui ressemble à un peptide cationique. Néanmoins il est nécessaire d'analyser la réponse à des peptides cationiques, produits par le système immunitaire. En effet, chez *S. pneumoniae* la réponse diffère selon les peptides antimicrobiens testés [336]. Nous pourrions alors comparer les réponses observées chez ces deux bactéries et mieux comprendre la réponse de *S. agalactiae* à la production de peptides cationiques par l'hôte au cours de l'infection.



**Tableau 6: Gènes différentiellement exprimés en conditions d'adaptation à l'ajout de colistine**

locus	Ratio TH Colistine/ TH	gène	Annotation
gbs0016	<b>0,49</b>	<i>pcsB</i>	pcsB protein
gbs0024	<b>2,65</b>	-	phosphoribosylformylglycinamide synthase
gbs0025	<b>2,25</b>	<i>purF</i>	amidophosphoribosyltransferase
gbs0026	<b>2,74</b>	<i>purM</i>	phosphoribosylformylglycinamide cyclo-ligase
gbs0027	<b>2,38</b>	<i>purN</i>	phosphoribosylglycinamide formyltransferase
gbs0028	<b>2,05</b>	-	acetyltransferase, GNAT family
gbs0031	<b>0,38</b>	-	group B streptococcal surface immunogenic protein
gbs0043	<b>2,06</b>	<i>purE</i>	phosphoribosylaminoimidazole carboxylase
gbs0048	<b>2,33</b>	-	transcriptional regulator, Cro/C1 family
gbs0060	<b>0,49</b>	<i>rplW</i>	ribosomal protein L23
gbs0061	<b>0,46</b>	<i>rplB</i>	ribosomal protein L2
gbs0062	<b>0,39</b>	<i>rpsS</i>	ribosomal protein S19
gbs0063	<b>0,34</b>	<i>rplV</i>	ribosomal protein L22
gbs0064	<b>0,40</b>	<i>rpsC</i>	ribosomal protein S3
gbs0065	<b>0,37</b>	<i>rplP</i>	ribosomal protein L16
gbs0066	<b>0,34</b>	<i>rpmC</i>	ribosomal protein L29
gbs0067	<b>0,36</b>	<i>rpsQ</i>	ribosomal protein S17
gbs0068	<b>0,39</b>	<i>rplN</i>	ribosomal protein L14
gbs0069	<b>0,35</b>	<i>rplX</i>	ribosomal protein L24
gbs0070	<b>0,39</b>	<i>rplE</i>	ribosomal protein L5
gbs0072	<b>0,44</b>	<i>rpsH</i>	ribosomal protein S8
gbs0082	<b>0,40</b>	<i>rpsM</i>	ribosomal protein S13
gbs0083	<b>0,41</b>	<i>rpsK</i>	ribosomal protein S11
gbs0084	<b>0,39</b>	<i>rpoA</i>	DNA-directed RNA polymerase, alpha subunit
gbs0108	<b>2,22</b>	-	deoxyuridine 5'-triphosphate nucleotidohydrolase
gbs0111	<b>0,45</b>	-	pyridine nucleotide-disulphide oxidoreductase protein
gbs0127	<b>2,09</b>	<i>rpmV</i>	ribosomal protein L28
gbs0142	<b>2,41</b>	-	penicillin-binding protein 4, putative
gbs0158	<b>2,45</b>	-	conserved hypothetical protein
gbs0196	<b>2,00</b>	-	oxidoreductase, putative
gbs0201	<b>0,46</b>	-	lipoprotein, putative
gbs0228	<b>0,39</b>	-	hypothetical protein
gbs0229	<b>0,44</b>	-	hypothetical protein
gbs0240	<b>2,05</b>	-	protein of unknown function
gbs0241	<b>2,43</b>	-	FtsK/SpoIIIE family protein
gbs0242	<b>2,80</b>	-	protein of unknown function
gbs0243	<b>2,86</b>	-	conserved hypothetical protein
gbs0253	<b>2,13</b>	-	hypothetical protein
gbs0254	<b>2,09</b>	-	conserved hypothetical protein
gbs0257	<b>6,54</b>	-	conserved hypothetical protein
gbs0272	<b>2,16</b>	-	PTS system, IIBC components
gbs0300	<b>2,21</b>	-	conserved hypothetical protein
gbs0344	<b>2,06</b>	-	membrane protein, putative
gbs0358	<b>3,10</b>	-	hypothetical protein
gbs0395	<b>2,22</b>	-	protein of unknown function
gbs0415	<b>0,45</b>	-	conserved hypothetical protein
gbs0419	<b>3,33</b>	-	protein of unknown function/lipoprotein
gbs0425	<b>0,50</b>	<i>polA</i>	DNA polymerase I
gbs0426	<b>0,41</b>	-	CoA-binding domain protein
gbs0428	<b>7,93</b>	-	cell wall surface anchor family protein
gbs0456	<b>0,39</b>	-	cell wall surface anchor family protein
gbs0459	<b>0,50</b>	-	hypothetical protein
gbs0470	<b>0,49</b>	<i>alp2</i>	surface protein Rib
gbs0495	<b>2,19</b>	-	magnesium transporter, CorA family
gbs0515	<b>2,30</b>	<i>nth</i>	endonuclease III
gbs0523	<b>0,46</b>	<i>murG</i>	UDP-N-acetylglucosamine--N-acetylmuramyl- transferase
gbs0524	<b>0,39</b>	<i>divIB</i>	cell division protein
gbs0526	<b>0,37</b>	<i>ftsZ</i>	cell division protein

gbs0527	0,29	-	ylmE protein, putative
gbs0528	0,38	-	ylmF protein
gbs0529	0,37	-	YGGT family protein
gbs0530	0,47	-	ylmH protein
gbs0546	0,47	<i>argR1</i>	arginine repressor ArgR, putative
gbs0547	0,42	<i>recN</i>	DNA repair protein RecN
gbs0552	2,21	-	hypothetical protein
gbs0553	5,29	<i>pyrD</i>	dihydroorotate dehydrogenase A
gbs0561	2,20	-	conserved hypothetical protein
gbs0583	3,04	-	adenosine deaminase, putative
gbs0594	0,45	-	transmembrane protein Vexp1
gbs0595	0,49	-	ABC transporter, ATP-binding protein Vexp2
gbs0600	2,03	-	conserved hypothetical protein
gbs0613	0,48	-	RNA methyltransferase, TrmA family
gbs0614	0,45	-	hypothetical protein
gbs0615	0,42	-	acid phosphatase, class B
gbs0661	3,88	-	DNA-entry nuclease, putative
gbs0685	2,07	-	DNA-binding response regulator
gbs0759	0,45	-	conserved hypothetical protein
gbs0785	0,48	<i>pbp2b</i>	penicillin-binding protein 2b
gbs0789	2,61	-	oxalate:formate antiporter
gbs0790	2,48	-	membrane protein, putative
gbs0805	2,11	-	conserved hypothetical protein
gbs0819	3,60	<i>nagB</i>	glucosamine-6-phosphate isomerase
gbs0825	0,41	-	hydrolase, haloacid dehalogenase-like family
gbs0827	0,37	<i>prsA</i>	protease maturation protein, putative
gbs0831	2,32	-	glycosyl transferase, family 8
gbs0837	0,42	<i>nrdE.2</i>	ribonucleoside-diphosphate reductase 2
gbs0847	2,36	-	conserved hypothetical protein
gbs0852	2,22	-	hypothetical protein
gbs0866	2,14	-	GtrA family protein
gbs0872	0,49	<i>glgC</i>	glucose-1-phosphate adenylyltransferase , subunit
gbs0873	0,47	<i>glgD</i>	glucose-1-phosphate adenylyltransferase, subunit
gbs0876	0,47	<i>atpB</i>	ATP synthase F0, A subunit
gbs0877	0,34	<i>atpF</i>	ATP synthase F0, B subunit
gbs0878	0,35	<i>atpH</i>	ATP synthase F1, delta subunit
gbs0890	0,41	<i>rexB</i>	exonuclease RexB
gbs0891	0,46	<i>rexA</i>	exonuclease RexA
gbs0897	0,40	<i>acoC</i>	acetoin dehydrogenase, thymine PPI dependent, E2 component, dihydroliipoamide acetyltransferase
gbs0911	0,48	<i>csn1</i>	protein of unknown function
gbs0916	2,32	<i>ndk</i>	nucleoside diphosphate kinase
gbs0933	0,31	<i>glmS</i>	glucosamine--fructose-6-phosphate aminotransferase
gbs0934	2,22	-	phnA protein
gbs0951	0,48	-	conserved hypothetical protein
gbs0952	2,06	-	Na <sup>+</sup> /H <sup>+</sup> exchanger family protein
gbs1021	0,37	<i>pepN</i>	aminopeptidase N
gbs1022	0,43	<i>phoU</i>	phosphate transport system regulatory protein PhoU
gbs1023	0,48	<i>pstB</i>	phosphate ABC transporter, ATP-binding protein
gbs1037	2,92	-	protease, putative
gbs1049	2,49	-	conserved hypothetical protein
gbs1050	2,38	-	carbon starvation protein CstA, putative
gbs1055	0,46	-	hypothetical protein
gbs1056	0,40	-	hypothetical protein
gbs1057	0,41	-	lipoprotein, putative
gbs1063	0,42	-	hypothetical protein
gbs1072	0,50	-	hypothetical protein
gbs1105	0,50	-	conserved hypothetical protein
gbs1111	3,97	-	4-oxalocrotonate tautomerase
gbs1141	2,13	-	membrane protein of unknown function
gbs1154	2,26	<i>guaC</i>	guanosine monophosphate reductase
gbs1156	3,12	-	conserved hypothetical protein
gbs1183	0,45	<i>folE</i>	GTP cyclohydrolase I

gbs1184	<b>0,50</b>	<i>folC.1</i>	folylpolyglutamate synthase
gbs1200	<b>0,44</b>	-	conserved hypothetical protein
gbs1202	<b>0,24</b>	-	gls24 protein, putative
gbs1203	<b>0,19</b>	-	conserved hypothetical protein
gbs1204	<b>0,16</b>	-	gls24 protein, putative
gbs1205	<b>0,19</b>	-	conserved hypothetical protein
gbs1206	<b>0,18</b>	-	conserved hypothetical protein
gbs1207	<b>0,38</b>	-	conserved hypothetical protein
gbs1208	<b>0,40</b>	-	conserved hypothetical protein
gbs1230	<b>2,03</b>	-	membrane protein, putative
gbs1236	<b>0,50</b>	<i>neuB</i>	N-acetyl neuramic acid synthetase
gbs1237.1	<b>0,50</b>	<i>cpsL</i>	protein of unknown function
gbs1239	<b>0,43</b>	<i>cpsJ</i>	glycosyl transferase CpsJ(V)
gbs1251	<b>0,45</b>	<i>deoD2</i>	purine nucleoside phosphorylase
gbs1252	<b>0,47</b>	-	voltage-gated chloride channel family protein
gbs1274	<b>0,50</b>	-	iminodiacetate oxidase, putative
gbs1295	<b>2,32</b>	-	conserved hypothetical protein
gbs1300	<b>2,36</b>	-	ISSdy1, transposase OrfA
gbs1301	<b>2,28</b>	-	ISSdy1, transposase OrfB
gbs1310	<b>2,35</b>	-	Transposase
gbs1312	<b>0,48</b>	-	DNA polymerase III alpha subunit (modular protein)
gbs1342	<b>0,50</b>	-	protein of unknown function
gbs1348	<b>2,12</b>	-	conserved hypothetical protein
gbs1349	<b>2,41</b>	-	hypothetical protein
gbs1350	<b>2,86</b>	-	hypothetical protein
gbs1351	<b>3,71</b>	-	conserved domain protein
gbs1377	<b>2,50</b>	<i>mmuM</i>	homocysteine S-methyltransferase
gbs1378	<b>2,49</b>	-	amino acid permease
gbs1379	<b>2,69</b>	-	hypothetical protein
gbs1380	<b>2,34</b>	-	hypothetical protein
gbs1415	<b>2,15</b>	-	conserved hypothetical protein
gbs1417	<b>0,39</b>	<i>fruB</i>	1-phosphofructokinase
gbs1457	<b>3,06</b>	-	ferredoxin, 4Fe-4S
gbs1465	<b>2,17</b>	<i>fruG</i>	iron compound ABC transporter, permease protein
gbs1478	<b>0,46</b>	<i>pilA</i>	cell wall surface anchor family protein
gbs1481	<b>0,49</b>	-	glycosyl transferase, group 2 family protein
gbs1484	<b>0,41</b>	-	glycosyl transferase, group 2 family protein
gbs1485	<b>0,47</b>	-	glycosyl transferase, group 2 family protein
gbs1486	<b>0,48</b>	-	nucleotide sugar dehydratase, putative
gbs1487	<b>0,47</b>	-	nucleotidyl transferase, putative
gbs1489	<b>0,49</b>	-	lipoprotein, putative
gbs1499	<b>2,17</b>	<i>rpsU</i>	ribosomal protein S21
gbs1505	<b>2,07</b>	-	conserved hypothetical protein
gbs1518	<b>0,49</b>	-	preprotein translocase SecA subunit, putative
gbs1535	<b>2,45</b>	-	conserved hypothetical protein
gbs1549	<b>2,18</b>	<i>secG</i>	preprotein translocase, SecG subunit
gbs1565	<b>2,10</b>	-	hypothetical protein
gbs1582	<b>2,15</b>	-	transcriptional regulator, LysR family
gbs1583	<b>2,35</b>	-	conserved hypothetical protein
gbs1608	<b>2,57</b>	-	hypothetical protein
gbs1609	<b>2,48</b>	-	hypothetical protein
gbs1621	<b>2,04</b>	<i>serC</i>	3-phosphoserine/phosphohydroxythreonine aminotransferase
gbs1633	<b>2,21</b>	-	conserved hypothetical protein
gbs1644	<b>0,46</b>	<i>scpA</i>	conserved hypothetical protein
gbs1645	<b>0,44</b>	-	integrase domain protein
gbs1663	<b>0,47</b>	-	acetyltransferase, GNAT family
gbs1695	<b>0,44</b>	-	dihydroxyacetone kinase family protein
gbs1696	<b>0,46</b>	-	conserved hypothetical protein
gbs1701	<b>0,45</b>	-	hypothetical protein
gbs1707	<b>0,42</b>	-	conserved hypothetical protein
gbs1766	<b>2,19</b>	-	conserved hypothetical protein
gbs1770	<b>2,05</b>	-	conserved hypothetical protein
gbs1776	<b>2,60</b>	-	membrane protein, putative

gbs1790	10,31	-	hypothetical protein
gbs1798	3,44	<i>rpsN2</i>	ribosomal protein S14
gbs1816	0,47	<i>cbf</i>	HD domain protein
gbs1828	0,44	-	hypothetical protein
gbs1836	2,02	<i>rpmH</i>	ribosomal protein L34
gbs1838	0,19	<i>opuABC</i>	amino acid ABC transporter, permease protein
gbs1839	0,25	<i>opuAA</i>	amino acid ABC transporter, ATP-binding protein
gbs1849	2,54	-	sugar-binding transcriptional regulator, LacI family
gbs1858	2,01	-	hypothetical protein
gbs1875	0,41	<i>ahpF</i>	alkyl hydroperoxide reductase, subunit F
gbs1881	2,47	-	membrane protein, putative
gbs1904	2,05	<i>yajC</i>	preprotein translocase, YajC subunit
gbs1907	2,34	<i>malP</i>	citrate carrier protein, CCS family
gbs1919	2,93	-	neuraminidase-related protein
gbs1940	2,78	-	membrane protein, putative
gbs1941	2,17	-	membrane protein, putative
gbs1944	2,24	<i>fasA</i>	response regulator
gbs1963	2,11	-	acetyltransferase, GNAT family
gbs1969	2,18	-	protein of unknown function
gbs1971	2,61	-	hypothetical protein
gbs1973	0,44	-	protein of unknown function
gbs1995	2,03	-	PAP2 family protein
gbs1996	2,79	-	conserved hypothetical protein
gbs1997	2,72	-	conserved hypothetical protein TIGR00730
gbs1998	3,08	-	protease, putative
gbs1999	0,26	-	rhodanese-like domain protein
gbs2000	0,28	<i>cfa</i>	cAMP factor
gbs2053	2,47	<i>csp</i>	cold shock protein, CSD family
gbs2069	2,11	-	Tn916, FtsK/SpoIIIE family protein
gbs2076	0,49	-	protein of unknown function
gbs2091	2,00	-	conserved hypothetical protein
gbs2102	2,35	-	membrane protein, putative
gbs2103	2,05	<i>trmU</i>	tRNA(5-methylaminomethyl-2-thiouridylate)- methyltransferase
gbs2106	0,22	-	protein of unknown function/lipoprotein
gbs2107	0,24	-	LysM domain protein
gbs2128	2,37	-	membrane protein, putative

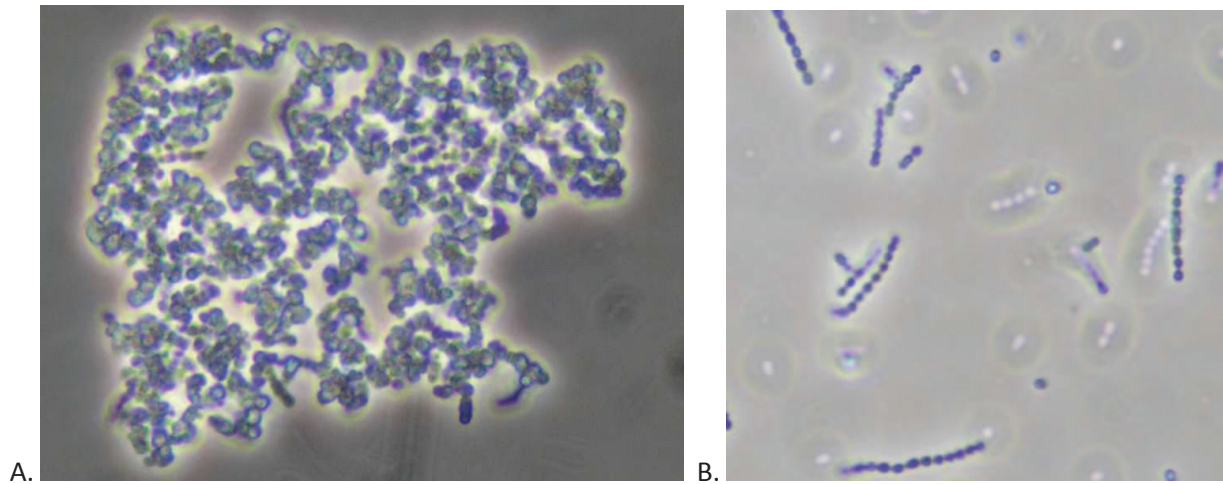
### (b) Interruption de *gbs1451* (*ltaS*)

En complément de l'étude de la modification du LTA, je souhaitais analyser l'impact de l'absence du LTA. Chez *S. aureus* il a été montré que la synthèse du LTA est dépendante du gène *ltaS* qui code pour une glycérol-phosphate polymérase [134]. Cette protéine est constituée de 5 segments transmembranaires et d'un grand domaine C-terminal extracellulaire. Le gène *gbs1451* de la souche NEM316 code une protéine qui présente 41% d'identité avec la protéine LtaS de *S. aureus*. Afin d'étudier les effets de l'absence de LTA au niveau de la paroi de *S. agalactiae*, j'ai tenté d'inactiver ce gène par simple recombinaison, après intégration d'un dérivé du plasmide à répllication thermosensible pG1, portant un fragment interne au gène *gbs1451*. Le fait de ne pas avoir obtenu d'intégration semble indiquer que ce gène est essentiel à la croissance de *S. agalactiae*. Le gène *gbs1451* est monocistronique, et la létalité ne doit pas être due à un effet

polaire. Chez *S. aureus*, le gène *ltaS* avait été initialement décrit comme essentiel [134]. Il a par la suite été montré que le mutant *ltaS* est thermosensible, et a une croissance lente à 30°C [135]. Chez *S. agalactiae*, l'insertion de pG1 nécessite une sélection à 40°C qui est sans doute incompatible avec la viabilité du mutant d'insertion.

### **b) Interruption de l'opéron *gbs1494-80* codant pour les fonctions nécessaires à la synthèse de l'antigène B.**

Très peu d'études ont été réalisées sur l'antigène B, un composé de la paroi de *S. agalactiae*. L'antigène B est un polysaccharide pariétal dont la structure est connue, mais dont la fonction biologique reste très hypothétique. Une souche mutante qui serait dépourvue d'antigène B a été décrite mais n'a pas été caractérisée [337]. Cette souche a une croissance défectueuse et forme de longues chaînes en conséquence d'un défaut de séparation des cellules. L'annotation du génome de la souche NEM316 [125] et une analyse bioinformatique récente, plus approfondie [120], ont permis de proposer que l'opéron *gbs1494 (rmlD) - gbs1480* était responsable de la biosynthèse de l'antigène B. Afin de caractériser la fonction de ce polymère, j'ai construit un mutant par insertion du plasmide thermosensible pG1 au niveau du second gène de cet opéron (*gbs1493*). Ce mutant présente une croissance très altérée ainsi que des défauts au niveau de la division cellulaire, visibles en microscopie optique (Figure 18). Ce résultat est en accord avec les données de Pincus et al. [337], et confirme les prédictions *in silico* sur le rôle de cet opéron dans la synthèse de l'antigène B [120]. Il confirme également l'importance de l'antigène B pour la structure et l'organisation de la paroi de *S. agalactiae*. L'analyse de nos données de transcriptome n'a pas mise en évidence de modulations de l'expression du locus codant pour l'antigène B, suggérant que l'antigène B est un composant structurel de la paroi.



**Figure 18: Photographie en microscopie optique du mutant insertionnel de l'antigène B et de la souche sauvage.** Le mutant insertionnel du gène *gbs1493* correspond à l'image A, et la souche NEM316 sauvage correspond à l'image B.

## 2. Réseaux de régulation impliquant des régulateurs de type RofA (publication n°2)

Chez *S. agalactiae* différents gènes, codant pour des protéines de surface impliquées dans l'interaction avec l'hôte, sont localisés à proximité des gènes codant pour les facteurs de transcription qui régulent leur expression. Nous avons voulu déterminer si ces facteurs de transcription pouvaient agir également sur l'expression d'autres composants de la paroi. Je me suis focalisée sur deux régulateurs de la famille RofA-like, Rga associé au locus *secA2* codant pour la protéine Srr1, et RogB associé à l'opéron *gbs1478-1474* codant pour le pilus PI-2a. Ces deux facteurs de transcription sont codés au niveau d'îlots génomiques, et nous voulions également déterminer ils étaient intégrés au sein des réseaux de régulation de *S. agalactiae*

Le facteur transcriptionnel RofA a été pour la première fois identifié chez *S. pyogenes* [338], où il agit comme activateur transcriptionnel du gène *prtF* codant pour une protéine de liaison à la fibronectine. Toujours chez *S. pyogenes*, il a été montré que les régulateurs RofA et Nra (tous deux de la famille RofA), sont capables d'agir comme activateurs transcriptionnels et répresseurs sur les mêmes cibles [339, 340]. Par exemple Nra joue respectivement le rôle de répresseur et d'activateur transcriptionnel sur les gènes du pilus [339, 341] dans les souches de sérotypes M49 et M53.

## **(1) Résultats**

### **(a) Rga et RogB régulent conjointement la synthèse du pilus PI-2a et de la protéine Srr.**

Afin de déterminer les rôles respectifs des deux régulateurs Rga et RogB, de type RofA, au sein des réseaux de régulation de *S. agalactiae*, j'ai étudié l'effet transcriptomique de la délétion des gènes correspondants, *rga* et *rogB*, dans un fond NEM316. Dans les mutants  $\Delta rogB$  et  $\Delta rga$ , respectivement 178 gènes et 96 gènes sont différentiellement exprimés par rapport à la souche sauvage. De manière intéressante l'expression du gène *Srr1* est induite dans le mutant  $\Delta rogB$  mais réprimée dans le mutant  $\Delta rga$ . Au contraire, l'expression de l'opéron PI-2a (*gbs1478-1474*) est induite dans le mutant  $\Delta rga$ , mais réprimée dans le mutant  $\Delta rogB$ . Nous avons montré l'existence d'une régulation croisée entre les deux régulateurs RogB et Rga, Rga activant l'expression de *rogB*. Au contraire Rga semble réprimer sa propre expression, une augmentation d'expression d'un facteur 9 étant observée au niveau de la partie non délétée du gène dans le mutant  $\Delta rga$ . L'interaction entre les deux systèmes de régulations laisse penser qu'il existe une relation fonctionnelle entre la protéine Srr1 et le pilus PI-2a. Cette relation fonctionnelle a été analysée par Shaynoor Dramsi (Article 2).

**Tableau 7: Ratio observé pour les gènes du locus pilus PI-2a, *srr1*, *rogB*, *rga*, et *gbs1605-06* dans les analyses transcriptomiques des mutants  $\Delta rga$ ,  $\Delta rogB$ ,  $\Delta CovRS$**

	Ratio $\Delta rga/WT$	Ratio $\Delta rogB/WT$	Ratio $\Delta covRS/WT$	
<b>gbs1474</b>	<b>0.09</b>	<b>0.98</b>	<b>3.91</b>	<i>pilC</i> , pilus anchoring subunit (LPXTG)
<b>gbs1475</b>	<b>0.11</b>	<b>0.69</b>	<b>3.88</b>	<i>srtC4</i> , pilus associated class C sortase
<b>gbs1476</b>	<b>0.09</b>	<b>0.69</b>	<b>3.57</b>	<i>srtC3</i> , pilus associated class C sortase
<b>gbs1477</b>	<b>0.03</b>	<b>0.4</b>	<b>3.81</b>	<i>pilB</i> , major pilin subunit (LPXTG)
<b>gbs1478</b>	<b>0.07</b>	<b>0.4</b>	<b>3.36</b>	<i>pilA</i> , pilus associated adhesin (LPXTG)
<b>gbs1479</b> <b>(<i>rogB</i>)</b>	<b>0.31</b>	<b>0.3</b>	<b>0.88</b>	<i>rogB</i> , transcriptional regulator RALP family
<b>gbs1529</b> <b>(<i>srr1</i>)</b>	<b>0.04</b>	<b>3.09</b>	<b>4.97</b>	<i>srr1</i> , Ser rich repeat glycoprotein (LPXTG)
<b>gbs1530</b> <b>(<i>rga</i>)</b>	<b>9*</b>	<b>1.56</b>	<b>3.71</b>	<i>rga</i> , transcriptional regulator RALP family
<b>gbs1605</b>	<b>0.44</b>	<b>0.53</b>	<b>101.14</b>	Glycosyltransferase, family 2
<b>gbs1606</b>	<b>0.43</b>	<b>0.48</b>	<b>77.63</b>	Glycosyltransferase, family 2

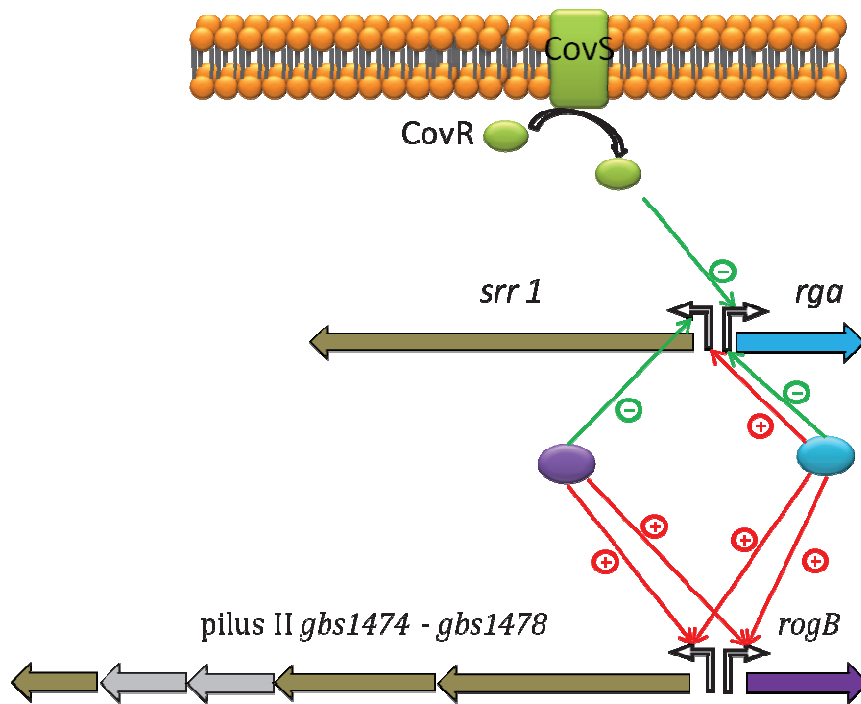
le ratio indiqué par une \* correspond au ratio observé pour l'unique sonde correspondant à la partie non délétée lors de la construction du mutant *rga*. Les ratios supérieurs à des effets de 2 sont indiqués en couleur.

#### (b) Une partie de la régulation génique par CovR dépend de Rga

Une analyse précédente utilisant des macroarrays montrait que l'expression de *rga* et non de *rogB* était induite d'un facteur 2 dans le mutant NEM316 $\Delta covSR$  [183], suggérant que CovR régulate l'expression de *rga*. J'ai confirmé que *rga* fait partie du régulon CovRS en étudiant l'effet de la mutation  $\Delta covSR$  à l'aide des microarrays Agilent qui donnent des résultats beaucoup plus précis. Les données du transcriptome indiquent que l'expression de *rga* est induite d'un facteur 3,7 et confirment que celle de *rogB* n'est pas affectée par la mutation de *covR*. L'analyse comparative des transcriptomes des mutants *rga* et *covSR* montre que, outre *rga*, le gène *ssr1* et l'opéron pilus PI-2a sont induits d'un facteur 5 et 3,8 respectivement dans le mutant  $\Delta covSR$ . Par ailleurs deux gènes codant pour des glycosyl transférases hypothétiques (*gbs1605-*




1606) sont également induits dans le mutant  $\Delta covSR$  et réprimés dans le mutant *rga*. Ceci indique qu'une partie de la réponse observée dans le mutant  $\Delta covSR$  fait intervenir Rga. CovR agirait comme répresseur de l'expression de *rga*, Rga étant lui même un activateur transcriptionnel de *srr1* et de *PI-2a*. Le site de fixation de CovR correspond à une région riche en AT. L'analyse de la séquence en amont du TSS du gène *rga* montre la présence d'un motif répété de 8 nt séparé de 19 nt qui pourrait correspondre au site de fixation de CovR (Figure 20). Cette figure montre que les sites de liaison entre CovR et Rga sont très proches, et suggère qu'il pourrait y avoir une compétition entre ces deux régulateurs au niveau de la région promotrice du gène *rga*. L'ensemble de ces données conduit à proposer le modèle de régulation décrit sur la Figure 19.



**Figure 19: Modèle d'action de Rga au sein des réseaux de régulation de *S. agalactiae***

Les protéines correspondant aux régulateurs transcriptionnels CovR, Rga et RogB sont respectivement indiquées par des ronds vert, bleu et violet. Les flèches rouges et vertes indiquent que ces régulateurs agissent respectivement comme activateurs transcriptionnel et répresseurs sur les gènes cibles.

ACCTTTATGAGGTTATCACTAATTATGACTAAAATAAATATACTTTAACCCTCATT  
 TGGAAATACTC**CAATAG**TGATTAATACTGATT**TTATTT**TATATGAAATT**GGGGAGTAAA**  
 (srr1)   
 CCCAAAAGGTATTTAAAATCTTTATTTTTTGAGAATTATT**ATTTTATA**AGGTAAAAT  
**GGTTTTT**CCATA**AAATTTTAGAAA**TAAAA**ACTCTTAATA**TAAAATATTC**CAATTTA**  
 AT**TTGAA**C**ATTTTATA**TTATTCCT**TATTAT**AGTATTA**ATG**  (rga)  
 TAAACTTTGTAAAATATAATAAGGAATAATATCATAATTAC

**Figure 20: Région promotrice des gènes divergents *srr1* et *rga***

Les séquences des boîtes -35 et -10 sont indiquées en gras et rouge, et le +1 de transcription est indiqué en violet pour les deux gènes. Les sites potentiels de fixation de CovR sont indiqués en bleu et les séquences protégées par Rga dans l'expérience d'empreinte à la DNaseI décrit dans l'article n°2 sont en vert.

## Publication n°2

---

Microbial Drug Resistance: <http://mc.manuscriptcentral.com/mdr>

**Rga, a RofA-like regulator, is the major transcriptional activator of the PI-2a pilus in *Streptococcus agalactiae***

Journal:	<i>Microbial Drug Resistance</i>
Manuscript ID:	Draft
Manuscript Type:	Great Wall Symposium
Date Submitted by the Author:	n/a
Complete List of Authors:	Dramsi, Shaynoor; Institut Pasteur, Microbiology Dubrac, Sarah; Institut Pasteur, Microbiology Konto-Ghiorghi, Yoan; Institut Pasteur, Microbiology Da Cunha, Violette; Institut Pasteur, Couv�, Elisabeth Glaser, Philippe; Institut Pasteur, Caliot, Elise D�barbouill�, Michel; Institut Pasteur, Microbiology Bellais, Samuel; Institut Cochin, Trieu-Cuot, Patrick; Institut Pasteur, Microbiology Mistou, Michel-Yves; INRA, UMR 1319 Micalis
Keyword:	Pathogenic bacteria
Abstract:	Rapid adaptation to changing environments is key in determining the outcome of infections caused by the opportunistic human pathogen <i>Streptococcus agalactiae</i> . We previously demonstrated that the RofA-like protein (RALP) regulators RogB and Rga activate their downstream divergently transcribed genes, i.e. the pilus operon PI-2a and the serine-rich repeat encoding gene <i>srr1</i> , respectively. Characterization of the Rga regulon by microarray revealed that the PI-2a pilus was strongly controlled by Rga, a result confirmed at the protein level. Complementation experiments showed that expression of Rga, but not RogB, in the double $\Delta$ rogB/ $\Delta$ rga mutant, or in the clinical strain 2603V/R displaying frameshift mutations in <i>rogB</i> and <i>rga</i> genes, is sufficient to restore wild-type expression levels of <i>Srr1</i> and PI-2a pilus. Biofilm formation was impaired in the $\Delta$ rga and $\Delta$ rga/ <i>rogB</i> mutants and restored upon complementation with <i>rga</i> . Paradoxically, adherence to intestinal epithelial cells was unchanged in the $\Delta$ rga mutant. Finally, the existence of several clinical isolates mutated in <i>rga</i> highlights the concept of strain specific regulatory networks.

1  
2  
3 **Rga, a RofA-like regulator, is the major transcriptional activator of the PI-**  
4 **2a pilus in *Streptococcus agalactiae***  
5  
6  
7  
8  
9  
10

11 **Shaynoor Dramsi<sup>1\*</sup>, Sarah Dubrac<sup>1</sup>, Yoan Konto-Ghiorghi<sup>1†</sup>, Violette Da Cunha<sup>2</sup>,**  
12 **Elisabeth Couvé<sup>2</sup>, Philippe Glaser<sup>2</sup>, Elise Caliot<sup>1</sup>, Michel Débarbouillé<sup>1</sup>, Samuel**  
13 **Bellais<sup>4,5,6</sup>, Patrick Trieu-Cuot<sup>1</sup> and Michel-Yves Mistou<sup>1,3</sup>**  
14  
15  
16  
17  
18  
19  
20  
21

22 <sup>1</sup>Institut Pasteur, Unité de Biologie des Bactéries Pathogènes à Gram-positif, Paris, F-75015

23 <sup>2</sup>Laboratoire Evolution et Génomique Bactériennes, Paris F-75015

24 <sup>3</sup>INRA, UMR 1319 Micalis, Jouy en Josas F-78350

25 <sup>4</sup>Institut Cochin, Université Paris Descartes Faculté de Médecine

26 <sup>5</sup>CNRS UMR 8104, Paris, France

27 <sup>6</sup>INSERM U1016, Paris, France

28 † Present address: Program in Epithelial Biology and Department of Dermatology, Stanford  
29 University School of Medicine, Stanford, California, USA  
30  
31  
32  
33  
34  
35  
36  
37

38 \* Author for correspondence: Shaynoor DRAMSI

39 Email: shaynoor.dramsi@pasteur.fr  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 Running title: Rga controls Srr1 and pili synthesis in GBS  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 1 Abstract

2 Rapid adaptation to changing environments is key in determining the outcome of  
3 infections caused by the opportunistic human pathogen *Streptococcus agalactiae*. We  
4 previously demonstrated that the RofA-like protein (RALP) regulators RogB and Rga activate  
5 their downstream divergently transcribed genes, *i.e.* the pilus operon PI-2a and the serine-rich  
6 repeat encoding gene *srr1*, respectively. Characterization of the Rga regulon by microarray  
7 revealed that the PI-2a pilus was strongly controlled by Rga, a result confirmed at the protein  
8 level. Complementation experiments showed that expression of Rga, but not RogB, in the  
9 double  $\Delta rogB/\Delta rga$  mutant, or in the clinical strain 2603V/R displaying frameshift mutations  
10 in *rogB* and *rga* genes, is sufficient to restore wild-type expression levels of Srr1 and PI-2a  
11 pilus. Biofilm formation was impaired in the  $\Delta rga$  and  $\Delta rga/rogB$  mutants and restored upon  
12 complementation with *rga*. Paradoxically, adherence to intestinal epithelial cells was  
13 unchanged in the  $\Delta rga$  mutant. Finally, the existence of several clinical isolates mutated in  
14 *rga* highlights the concept of strain specific regulatory networks.

15

## 1 Introduction

2  
3  
4  
5  
6  
7 *Streptococcus agalactiae* (also known as group B streptococcus, GBS) is a Gram-positive  
8 encapsulated commensal bacterium of the human intestine that is also present in the vagina of  
9 15-30% of healthy women <sup>2</sup>. However, in certain circumstances, GBS can become a life-  
10 threatening pathogen in neonates <sup>9</sup>. Two distinct GBS-associated clinical syndromes, referred  
11 to as early-onset disease (EOD) and late-onset disease (LOD) have been recognized in  
12 neonates depending on the age of the infant at the time of disease manifestations. For EOD,  
13 epidemiological studies have documented how commonly GBS are transmitted from "carrier"  
14 mothers to newborn infants <sup>38</sup>. Transmission occurs by inhalation of GBS-contaminated  
15 amniotic or vaginal fluid during delivery, followed by bacterial translocation across the  
16 respiratory epithelium that rapidly progresses into subsequent systemic infection. The lung is  
17 most likely the portal of entry. In contrast, for LOD, the mode of transmission and the  
18 infection route has remained elusive. LOD is characterized by bloodstream infection with a  
19 high risk of progression to meningitis and is strongly associated with GBS strains belonging  
20 to the "hypervirulent" clonal complex 17 <sup>16,17,29</sup>. A recent study strongly supports an intestinal  
21 portal of entry for LOD <sup>39</sup>. GBS is also an emerging pathogen of adults with elderly,  
22 immunocompromised and those with diabetes and malignancies being particularly at risk <sup>8,10</sup>.  
23 Clinical manifestations are varied and include skin, soft tissue and urinary tract infections,  
24 bacteremia, pneumonia, arthritis and endocarditis <sup>8,10</sup>.  
25 During the last decade, key virulence factors of GBS involved in colonization, survival,  
26 spread and persistence of the bacterium within the human host have been identified (for a  
27 recent review, see <sup>32</sup>). In addition, the availability of eight genome sequences accelerated our  
28 understanding of the pathogen's strategy <sup>40</sup>. Rapid adaptation to changing environments  
29 encountered during infection is essential for GBS survival that need to colonize various  
30 tissues, obtain nutrients, or evade the host immune response. The genome sequences of GBS  
31 strains have revealed the presence of 17-20 two-component systems (TCS) that can respond  
32 to changes in the external environment. One of the best-characterized TCS is the CovS/CovR  
33 system that plays a major role in virulence <sup>15,21</sup>. The regulatory activity of CovR is not only  
34 modulated by its cognate sensor CovS but also by a serine/threonine kinase Stk1 that  
35 inactivates CovR by phosphorylation on threonine residue 65 <sup>22</sup>.  
36 GBS NEM316 encodes 6 "stand-alone" transcriptional regulators <sup>11</sup> amongst which RogB,  
37 Rga and Gbs1426 belong to the RofA-like family (RALP) (for a review, see <sup>19</sup>). RogB and  
38 Rga are the closest relatives of this family exhibiting 47% identity whereas Gbs1426 is more

1  
2  
3 35 distantly related, displaying 23% and 24% identity to RogB and Rga respectively. RofA, the  
4  
5 36 founding member of this family, was first identified in Group A Streptococci (GAS) as a  
6  
7 37 positive transcriptional activator of the adjacent *prtF* gene encoding the fibronectin binding  
8  
9 38 protein SfbI<sup>30</sup>. Nra, the second RALP regulator identified in GAS, was found to be a negative  
10  
11 39 regulator for an adjacent gene that encodes the collagen-binding protein Cpa, as well as an  
12  
13 40 unlinked gene encoding the fibronectin-binding protein PrtF2<sup>30</sup>. RALP proteins exhibit  
14  
15 41 strong similarities with another global regulator of M protein synthesis in GAS called Mga  
16  
17 42 (Hondorp et al., 2007). Mga and RALP proteins are about 500 amino acids long and contain 3  
18  
19 43 conserved Pfam domains: i) HTH\_Mga (PF08280); ii) Mga (PF05043); and iii) PRD  
20  
21 44 (PF08270). Both Helix-Turn-Helix found in PF08280 and PF05043 were involved in DNA-  
22  
23 45 binding<sup>27</sup> (see Fig. S1).

24  
25 46 In the present work, we demonstrate the central role of Rga in controlling the synthesis of two  
26  
27 47 important GBS NEM316 surface adhesins, namely PI-2a pilus and Srr1. Surprisingly, the  
28  
29 48 adherence of a *rga* knock-out strain to epithelial cells was similar to that of the wild-type  
30  
31 49 (WT) strain suggesting that the display of surface proteins is a highly dynamic process which  
32  
33 50 is dependent on expression of multiple adhesins.  
34  
35 51



## 1 Results

### 3 Rga microarray analysis

4 To define the respective roles of the paralogous regulators Rga and RogB, we carried out a  
5 genome-wide transcriptional analysis of the WT NEM316 strain and its isogenic  $\Delta rga$  and  
6  $\Delta rogB$  mutant strains. In the  $\Delta rga$  strain, 60 genes were found down regulated and 36 up  
7 regulated compared to the parental NEM316 using a 2-fold cut off threshold. These regulated  
8 genes are often grouped into operons and encode proteins associated to a large variety of  
9 unrelated functions suggesting that Rga is a pleiotropic regulator. Among the down-regulated  
10 genes, we identified its cognate target gene *srr1* but, most importantly, the regulator-encoding  
11 gene *rogB* and its target operon *gbs1478-1474* (named PI-2a in <sup>35</sup>) (Table 1). In the  $\Delta rogB$   
12 strain, 96 genes were found down regulated and 82 genes up regulated. The down-regulated  
13 genes included the PI-2a operon and the two component regulatory systems *ciaRH* and *relPR*,  
14 whereas *srr1* was among the most up regulated genes. RogB, alike Rga, may impact on the  
15 expression of other regulatory genes making it difficult to identify genes which are directly  
16 regulated by either regulators from those which are indirectly regulated.

17 A previous macroarray analysis performed in our laboratory showed that *rga* (*gbs1530*), but  
18 not *rogB*, was 2-fold induced in the NEM316 $\Delta covSR$  mutant <sup>21</sup>. This observation was  
19 confirmed in this study by using microarray analysis (E-MEXP-3304). We thus compared the  
20 Rga regulon which included 96 differentially regulated genes (2-fold cut-off threshold) to the  
21 CovSR and RogB regulons using Cluster 3 and TreeView programs. The subset of 61 genes  
22 differentially regulated in at least two microarrays is shown in Fig. 1. The most striking group  
23 of genes likely regulated through a CovSR-dependent Rga regulation includes *srr1* and the  
24 PI-2a pilus operon (*gbs1478-1474*), the glycosyltransferase encoding genes *gbs1605-1606*,  
25 and that encoding the surface adhesin BibA (*gbs2018*) (Table 1).

26 Quantitative analysis of the results shows that the expression of the PI-2a operon is positively  
27 regulated by both Rga and RogB, but that the impact of the *rga* deletion is 10 times that of the  
28 *rogB* mutation.

### 30 Rga is the master regulator of the pilus operon

31 In order to determine the hierarchy between Rga and RogB in the transcriptional activation of  
32 the PI-2a pilus operon (*gbs1478-1474*) and *srr1* (*gbs1529*), quantitative RT-PCR experiments

1  
2  
3 33 on *srr1* and *pilB* were performed using the WT strain NEM316 and its isogenic  $\Delta rga$  and  
4  $\Delta rogB$  mutant strains. The results shown in Fig. 2 demonstrated that Rga is the major  
5 transcriptional activator of PI-2a pilus and *srr1* transcription (> 100 fold on *pilB*, the second  
6  
7 35 transcriptional activator of PI-2a pilus and *srr1* transcription (> 100 fold on *pilB*, the second  
8  
9 36 gene of the operon, encoding the major pilin). As previously described, RogB only slightly  
10  
11 37 activates *pilB* transcription by 2-3 fold<sup>6,13</sup>. We also observed a slight repression of *srr1*  
12  
13 38 transcription by RogB confirming the transcriptomic data.

14  
15 39 By dot-immunoblotting with specific polyclonal antibodies against Srr1 and PilB on  
16  
17 40 exponentially growing bacteria, we determined Srr1 and PilB protein levels in the WT strain  
18  
19 41 NEM316 and the two isogenic mutants  $\Delta rga$  and  $\Delta rogB$ . As a control for bacterial loading,  
20  
21 42 we used an antibody against formaldehyde killed GBS bacteria. No signal could be detected  
22  
23 43 in the  $\Delta rga$  mutant indicating that Rga is the major activator of both Srr1 and PilB synthesis  
24  
25 44 (Fig. 3). Trans-complementation with a plasmid-borne *rga* restored both Srr1 and PilB  
26  
27 45 synthesis (Fig. 3).

28  
29 46 The complex interaction between RogB and Rga prompted us to construct a double  
30  
31 47  $\Delta rga/rogB$  mutant derivative of strain NEM316. We then introduced plasmid borne *rga* or  
32  
33 48 *rogB* under the control of a constitutive promoter both in the WT and double mutant strains.  
34  
35 49 Western blot and dot-blot analyses showed that i) overexpression of *rogB* or *rga* in the WT  
36  
37 50 strain increased the formation of PilB polymers; and ii) expression of *rga* only, but not *rogB*,  
38  
39 51 in the  $\Delta rga/rogB$  strain was sufficient to restore WT levels of pilus synthesis (Fig. 4).  
40  
41 52 Similarly, overexpression of *rga* in the WT strain increased the expression level of Srr1 at the  
42  
43 53 cell surface and was sufficient to restore WT levels of Srr1 in the  $\Delta rga/rogB$  strain (data not  
44  
45 54 shown).

46  
47 55 Analysis of *rga* sequence in the three complete GBS genomes and 4 partially sequenced  
48  
49 56 genomes revealed that strains A909, 2603V/R, and CJB111 likely produced truncated Rga  
50  
51 57 resulting from frameshift mutation and/or mutation creating an internal stop codon (Fig. S1).  
52  
53 58 Interestingly, the GBS strain 2603V/R responsible for adult invasive disease also exhibited a  
54  
55 59 frameshift mutation in the *rogB* gene, thus constituting a natural *rga/rogB* double mutant. We  
56  
57 60 first showed by immunoblotting analyses that 2603V/R did not synthesize any detectable  
58  
59 61 levels of Srr1 and PilB proteins (not shown). By introducing in *trans* a functional copy of *rga*  
60  
61 62 (pRga) in strain 2603V/R, we restored expression of both PilB polymers and Srr1, whereas  
62  
63 63 the trans-expression of *rogB* did not (Fig. 4 for PilB, not shown for Srr1). Altogether these  
64  
65 64 results indicate that expression of *rga* is necessary and sufficient for both PI-2a pilus and Srr1  
66  
67 65 expression.

68  
69  
70 66

### 67 **Characterization of PI-2a and Srr1 promoters**

68 To gain further insight in the molecular mechanisms underlying the Rga-dependant  
69 regulation, transcription start sites were localized in the *rogB-pilA* and *rga-srr1* intergenic  
70 regions (Fig. S2). Single non-overlapping transcriptional start sites were found upstream *pilA*  
71 (the first gene of the PI-2a operon), *rogB*, *srr1* and *rga* genes, suggesting no transcriptional  
72 interference between the divergently transcribed genes (Fig. S2).

73 *In vitro* binding assays were performed to test whether the Rga-dependant activation results  
74 from direct interaction of Rga with its target promoters. Our many attempts to express and  
75 purify the recombinant Rga protein in *E. coli* using different systems (Rga-6His or MBP-Rga  
76 or RgA-intein) have failed. In all cases, the protein was either insoluble or degraded (data not  
77 shown). We therefore cloned the entire *rga* with a C-terminal histidyl tag in the multicopy  
78 vector pAT18 giving rise to pRga-6His. This plasmid, able to complement the  $\Delta$ *rga* mutant  
79 (not shown), was expressed in *L. lactis* strain NZ9000 under the control of a constitutive  
80 promoter. The recombinant protein Rga-6His was purified from *L. lactis* and its identity  
81 verified by mass spectrometry analysis (data not shown). DNase I footprinting assays were  
82 carried out using either the purified Rga-6His protein (20  $\mu$ g) or total protein extracts (25-100  
83  $\mu$ g) from *L. lactis* (pRga-6His) or *L. lactis* (pAT18). As shown in Fig. 5, two close DNA  
84 sequences in the *srr1* promoter region from -100 to -69 and from -64 to -45 relative to the  
85 transcriptional start site were protected with the total extracts of *L. lactis* (pRga-6His). No  
86 protection was detected with control *L. lactis* (pAT18) extracts, neither with the purified  
87 protein Rga-6His. The same experiment was done with the PI-2a promoter region and no  
88 protection against DNase I digestion was observed (data not shown). These results show that  
89 i) *srr1* is directly regulated by Rga while the PI-2a operon is probably indirectly regulated;  
90 and ii) that purification of Rga is detrimental to its DNA-binding activity, either because of  
91 protein instability or of the loss of a cytoplasmic cofactor.

92

### 93 **Biofilm formation is abolished in the $\Delta$ *rga* mutant and restored upon complementation**

94 A major role of GBS PI-2a pilus is its implication in biofilm formation<sup>18,34</sup>. Indeed,  
95 expression of the NEM316 PI-2a pilus operon in the non-pathogenic *L. lactis* strain NZ9000  
96 conferred to the recombinant bacteria the ability to form biofilm (Fig. S3 and<sup>34</sup>). We  
97 therefore tested the isogenic  $\Delta$ *rga* and  $\Delta$ *rogB* mutants for their ability to form biofilm on  
98 polystyrene plates using LB glucose medium. As shown in Fig. 6A, biofilm formation is  
99 slightly reduced in the  $\Delta$ *rogB* mutant but strongly impaired in the  $\Delta$ *rga* mutant as compared to  
100 the parental NEM316 strain. Complementation of  $\Delta$ *rogB* and  $\Delta$ *rga* by plasmid overexpression

1  
2  
3 101 of *rogB* and *rga* respectively restored biofilm formation (Fig. 6B). Finally, we showed that  
4 102 overexpression of *rga* in the  $\Delta$ *rga/rogB* was sufficient to confer to the strain the ability to  
5 103 form biofilm whereas overexpressing vector alone or *rogB* did not (Fig. 6C). Similarly,  
6 104 overexpression of *rga* in GBS strain 2603V/R conferred to this GBS clinical isolate the  
7 105 capacity to form biofilm in our experimental conditions. Finally, overexpression of *rga* in the  
8 106 WT strain increased the levels of biofilm as compared to the WT strain carrying the vector  
9 107 alone or the plasmid carrying *rogB* (Fig. 6C).

10 108 The second important role assigned to GBS pilus PI-2A is to promote adhesion to alveolar  
11 109 epithelial cells A549<sup>6,18</sup> whereas Srr1 was shown to be a surface adhesin<sup>28</sup>. We thus tested  
12 110 the adhesive capacity of the  $\Delta$ *rga* mutant to A549 alveolar cells and TC7 intestinal cells.  
13 111 Surprisingly, we did not find any significant difference for adherence to these cell lines  
14 112 between  $\Delta$ *rga* mutant and WT strain (Fig. 7 and data not shown). This led us to hypothesize  
15 113 that expression and/or display of other surface proteins in the  $\Delta$ *rga* mutant compensate for the  
16 114 absence of GBS adhesins Srr1 and PI-2a pilus.

## 1 Discussion

2 We show here that the PI-2a pilus and Srr1, two major surface components previously  
3 characterized in our laboratory, are under the positive control of a common transcriptional  
4 regulator Rga belonging to the RALP family. We demonstrated the pre-eminence of Rga over  
5 RogB, the second closest RALP paralog exhibiting 47 % identity with Rga and previously  
6 shown to activate the transcription of the pilus operon PI-2a<sup>6,13</sup>. It is worth noting that the  
7 synthesis of these adhesins is partially dependent upon the TCS CovSR which acts by  
8 negatively controlling the transcription of *rga*. A working model of regulation is depicted in  
9 Fig. 8.

10 Relatively little is known about the regulation of virulence factors in GBS besides the central  
11 role of CovSR, a two-component system (TCS). We and others previously showed that the  
12 response regulator CovR is a pleiotropic regulator of virulence genes which binds to the  
13 promoter regions of the genes that it controls<sup>15,21</sup>. A consensus CovR recognition sequence  
14 has been determined (5'-TATTTAAT-3') using both DNase I footprinting and  
15 computational analyses<sup>21</sup>. Interestingly, the *rga* gene was 2-fold induced in the  $\Delta covRS$   
16 strain, indicating that part of the response of *S. agalactiae* to the *covR* mutation may be  
17 indirect. By examining carefully the *rga* promoter region, we found a putative binding site for  
18 CovR located 40 bp from the transcriptional start site (TSS) with two mismatches in the  
19 recognition sequence.

20 Rga and RogB belong to the Mga superfamily which includes the RofA-like protein family  
21 or RALPs originally described in GAS (for a recent review, see<sup>25</sup>). These transcriptional  
22 regulators coined as “stand-alone” response regulators were previously shown to control  
23 global virulence regulons in response to changing environmental conditions<sup>14</sup>. Like the other  
24 members of this family of regulators, Rga is predicted to exhibit two successive helix-turn-  
25 helix motifs at its N-terminus, which is consistent with a DNA-binding function. Only two  
26 poorly defined consensus sites have been defined for Mga and RofA<sup>12,26</sup>. Interestingly the  
27 DNA sequence protected by Rga found upstream from *srr1* (boxes I and II) is reminiscent of  
28 the RofA-binding site found within the intergenic region shared by *rofA* and *prtF* at a similar  
29 position relative to the TSS<sup>12</sup>. Computational analyses of the various promoter regions of Rga  
30 target genes were performed but did not reveal any clear consensus sequence suggesting a  
31 complex regulation.

1  
2  
3 32  
4  
5  
6

33 During the preparation of this manuscript, another study reporting the role of Rga in pilus  
34 formation in *S. agalactiae* NEM316 was published<sup>37</sup>. These authors claim that recombinant  
35 Rga purified from *Escherichia coli* can bind to the promoter regions of *pilA* and *fbsA*. We  
36 were unable to see any convincing gel shift on *pilA* and *srr1* promoter regions with purified  
37 Rga under various conditions (not shown). Using the DNase I protection assay, a protected  
38 region in the *srr1* promoter region was defined with total extracts of *L. lactis* expressing *rga*  
39 compared to *L. lactis* carrying the vector alone. But no protection was observed with the  
40 recombinant purified Rga-6His alone. Collectively, our results Rga suggest that other  
41 bacterial components are necessary to promote binding of Rga to certain of its target  
42 promoters.

21  
22 43  
23

24 44 Genome-wide transcriptional analysis indicated that Rga not only controls the expression of  
25 genes encoding surface proteins but also additional aspects of bacterial physiology in  
26 particular arginine catabolism (*gbs2083-2085*; *gbs2122-2126*). Both *gbs2083-2085* and  
27 46 *gbs2122-2126* operons encode genes showing homologies to those found in the arginine  
28 deiminase pathway (ADI). This multi-enzyme complex catalyzes the conversion of arginine  
29 47 to ornithine, ammonium, and carbon dioxide with the concomitant production of ATP. ADI-  
30 48 expressing bacteria were shown to better overcome oxygen and nutrient starvation but also to  
31 better survive in acidic environments<sup>24</sup>. From the microarray data (Table 1), it appears that  
32 49 Rga strongly activates the ADI pathway. We therefore tested the survival of GBS NEM316  
33 and its isogenic mutant strain  $\Delta$ *rga* in various pH conditions (6.0, 5.5, 5 and 4). We found that  
34 50 the growth of the  $\Delta$ *rga* mutant was strongly impaired in acidic conditions, with a 2-log  
35 51 reduction in viability after 2 h at pH 4 compared to the WT strain (data not shown). It is thus  
36 52 tempting to speculate that Rga may play a major role during vaginal colonization of GBS  
37 53 strain NEM316 and that the pilus PI-2A and Srr1 are the major adhesins involved in this  
38 54 process.  
39 55  
40 56  
41 57  
42 58  
43 59  
44 60

45  
46  
47  
48  
49

50 60 We and others previously showed a major role of GBS PI-2a pilus in biofilm formation<sup>18,34</sup>.  
51 61 Here, we further strengthen this result by demonstrating a strict correlation between the PI-2a  
52 62 expression and the ability of the different mutants tested to form biofilm (Fig. 6 and S3). We  
53 63 previously showed that both PI-2a pilus and the surface glycoprotein Srr1 are bacterial  
54 64 adhesins for host epithelial cells<sup>6,18,28</sup>. To our surprise, the  $\Delta$ *rga* mutant, which is strongly

55  
56  
57  
58  
59  
60

1  
2  
3 65 impaired for both PI-2a and Srr1 synthesis, was found to be as adherent as the parental strains  
4  
5 66 on human epithelial cells A549 and TC7. We suggest a profound remodelling of the bacterial  
6  
7 67 surface in the  $\Delta rga$  mutant with the exposition of other potential adhesins such as FbsA  
8  
9 68 (Samen et al., 2011), a major fibrinogen binding protein, that could potentially compensate  
10  
11 69 for the loss of PI-2a and Srr1. Global surface proteome changes might also explain the  
12  
13 70 behaviour of two other NEM316 derivatives,  $\Delta gbs1478-74$  and  $\Delta srtC3C4$  mutant in the PI-2a  
14  
15 71 locus, which were found to be as adherent as the WT strain to epithelial cells (S. Dramsi,  
16  
17 72 unpublished data). Overall these results suggest that adherence of GBS to epithelial cells is  
18  
19 73 probably mediated by multiple adhesins and that removal of one or several adhesins from the  
20  
21 74 GBS cell surface is often compensated by the display of other surface proteins involved in the  
22  
23 75 same function proving the robustness of the system.

24  
25 76 In contrast, the biofilm formation in LB glucose is strictly dependent on PI-2a pilus synthesis  
26  
27 77 in GBS. Moreover, the recombinant *L. lactis* strain overexpressing the PI-2a pilus operon are  
28  
29 78 able to form biofilm<sup>34</sup> (Fig. S3). We also noticed that overexpression of *rga* in the  $\Delta rga$  strain  
30  
31 79 which increased biofilm formation decreased the adherence level to epithelial cells suggesting  
32  
33 80 in these conditions a fine-tuned balance between bacteria-bacteria and bacteria-host  
34  
35 81 interactions (data not shown). Of note, the recombinant *L. lactis* strains overexpressing the PI-  
36  
37 82 2a pilus or the pilus-associated adhesin PilA were unable to adhere to epithelial cells as  
38  
39 83 compared to the control *L. lactis* strain carrying the empty plasmid (Fig. S3).

40  
41 84  
42  
43 85 Our study was performed in strain NEM316 but we noticed that, among the other available  
44  
45 86 GBS genomes, at least three strains (A909, CJB111, 2603V/R) display a frameshift mutation  
46  
47 87 in the *rga* gene at various positions leading to a truncated protein. Similarly, the GBS strains  
48  
49 88 2603V/R and 515 display a frameshift mutation in the *rogB* gene (Fig. S1). It is worth noting  
50  
51 89 that in the GBS strain COH1, which belongs to the hypervirulent clonal complex ST-17, the  
52  
53 90 chromosomal locus PI-2a is replaced by PI-2b and that the *srr1* locus is missing;  
54  
55 91 consequently, this strain does not express neither Rga nor RogB regulators<sup>1</sup>. These  
56  
57 92 observations reveal that GBS clinical isolates exhibit a diverse repertoire of surface  
58  
59 93 components and associated transcriptional regulators.

60  
61 94 This work leads us to consider the bacterial surface as a dynamic structure susceptible to  
62  
63 95 unpredictable remodeling that may have important functional consequences. Thus, the  
64  
65 96 concept of pathogenicity in streptococci appears more complex than previously thought.  
66  
67 97

## 1 **Materials and Methods**

### 3 *Bacterial strains, media and growth conditions*

4 Bacterial strains and plasmids used in this study are listed in Table S1. *Streptococcus*  
5 *agalactiae* NEM316 is a ST-23 serotype III strain whose genome has been sequenced<sup>11</sup>. The  
6 NEM316 $\Delta$ rga and NEM316 $\Delta$ rogB mutants were constructed previously<sup>6,28</sup>. The *Escherichia*  
7 *coli* DH5 $\alpha$  (Invitrogen) used for cloning experiments was grown in Luria-Bertani (LB)  
8 medium. Unless otherwise specified, *S. agalactiae* was cultured in Todd-Hewitt (TH) broth  
9 (standing filled flasks) or agar (Difco Laboratories, Detroit, MI). Erythromycin was used at  
10 150  $\mu$ g/ml for *E. coli* and 10  $\mu$ g/ml for *S. agalactiae*. *Lactococcus lactis* strain NZ9000<sup>20</sup> was  
11 grown in M17 medium supplemented with 1% glucose (GM17). Heterologous expression of  
12 PI-2a pilus in *L. lactis* strain was realized as follows: The five genes *gbs1478-1474* were  
13 amplified from NEM316 genomic DNA and cloned in the lactococcal vector pOri23, a high  
14 copy number erythromycin resistance plasmid expressing the cloned gene from the strong  
15 constitutive promoter P23<sup>31</sup>. The primers used for amplification of *gbs1478* (*pilA*), *gbs1477*  
16 (*pilB*), *gbs1474* (*pilC*) and the entire operon PI-2a (*gbs1478-1474*) are shown in Table S1.  
17 The various PCR products were then cut by the appropriate enzymes (New England Biolabs),  
18 purified and cloned into pOri23. The resulting plasmids were verified by DNA sequence and  
19 introduced into electrocompetent *L. lactis* NZ9000.

### 21 *General DNA techniques*

22 Standard recombinant techniques were used for nucleic acid cloning and restriction analysis  
23<sup>36</sup>. Plasmid DNA from *E. coli* was prepared by rapid alkaline lysis using the QIAprep Spin  
24 Miniprep Kit (Qiagen). Genomic DNA from *S. agalactiae* was prepared using the DNeasy  
25 Blood and Tissue Kit (Qiagen). PCR was carried out with Phusion Taq polymerase as  
26 described by the manufacturer (Finnzymes).

### 28 *Construction of mutants and complemented strains*

29 The double mutant  $\Delta$ rogB/ $\Delta$ rga was constructed as described for NEM316 $\Delta$ rogB but in the  
30 NEM316 $\Delta$ rga genetic background.  
31 The primer pairs RgaF-RgaR and RogBF-RogBR were used to amplify the sequence  
32 encoding the *rga* and *rogB* genes respectively (Table S1). The resulting PCR fragments were



1  
2  
3 33 digested with *Bam*H1-*Pst*1 and cloned into the shuttle vector pTCVErm-Ptet to generate pRga  
4 and pRogB respectively. The resulting plasmids were introduced by electroporation into the  
5 various strains.  
6  
7  
8 36

9  
10 37 *Expression and purification of recombinant Sag1407*

11  
12  
13 38 Sag1407 DNA coding region was amplified by PCR using genomic DNA of GBS 2603V/R  
14 as template and primers Sag1407*Nco* and Sag1407*Bam* (Table S1). The PCR product was  
15 digested with *Nco*I and *Bam*HI enzymes and cloned into pET2817 (Tazi et al., 2010). The  
16 resulting plasmid was introduced into *E. coli* BL21λDE3/pDIA17 for protein expression.  
17  
18 41 Recombinant 6xHis-SAG1407 protein was purified under native conditions by affinity  
19 chromatography on Ni-NTA column according to manufacturer's recommendation (Qiagen)  
20 followed by exclusion chromatography on Superdex 200 column (GE Healthcare). Protein  
21 purity was checked on SDS-PAGE and accurate protein concentration was determined using  
22 the Bradford protein assay. Rabbit polyclonal antibodies were produced and purchased from  
23 Covalab (Villeurbanne, France).  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

34 48  
35  
36  
37  
38  
39 49 *Microarray analysis*

40 We used oligo-arrays carrying three oligonucleotides per gene, that was demonstrated as a  
41 much more sensitive method than the previously used DNA macro-arrays<sup>21</sup>. RNA was  
42 reverse-transcribed with Superscript indirect cDNA kit (Invitrogen) and labeled with Cy5 or  
43 Cy3 (Amersham Biosciences) according to supplier's instructions. The microarray containing  
44 6835 60-mer oligonucleotides specific for 2134 predicted genes of the genome of NEM316  
45 strain and for all intergenic regions longer than 100 nucleotides has been designed using the  
46 program OligoArray (<http://berry.engin.umich.edu/oligoarray/>). Based on these sequences a  
47 custom oligonucleotide array was manufactured (Agilent Technologies) with a final density  
48 of 15K. For hybridization, Cy3 and Cy5 target quantities were normalized at 150 pmol.  
49  
50 Arrays were scanned in an Axon 4000B dual laser scanner connected to the GenePix Pro 6.0  
51 image acquisition and analysis software package. Slides are scanned at 635nm and 532nm  
52 wavelength (respectively for Cy5 and Cy3). An initial pre-scan of the slide is performed to  
53 determine success of hybridization and to allow defining the boundaries of the array features.  
54  
55 PMT gain settings are adjusted at this step to ensure that the overall red-to-green ratio is  
56 similar (PMT settings are 635nm = 460, 532nm = 425). Spots were excluded from analysis in  
57  
58  
59  
60

1  
2  
3 65 case of high local background fluorescence slide abnormalities or weak intensity. No  
4  
5 66 background was subtracted. For the purpose of a statistical analysis, RNA was prepared from  
6  
7 67 three independent cultures and each RNA sample was hybridized twice to the microarrays  
8  
9 68 (dye swap), leading to 6 values. Data normalization and differential analysis were conducted  
10  
11 69 using the R software (<http://www.r-project.org>). A global intensity-dependent normalization  
12  
13 70 using the LOESS procedure<sup>41</sup> was performed on a slide-by-slide basis (BioConductor  
14  
15 71 package marray; <http://www.bioconductor.org/packages/bioc/html/marray.html>) to correct the  
16  
17 72 dye bias. To determine differentially expressed genes, we performed a paired t test using the  
18  
19 73 VM method (VarMixt package<sup>4</sup> together with the Benjamini and Yekutieli p-value correction  
20  
21 74 method<sup>33</sup>. For each gene 3 probes were present on the microarray. The cut-off for the  
22  
23 75 expression ratio was set to either superior/equal to 2 or inferior/equal to 0.5 and the general  
24  
25 76 ratio of expression of each gene was calculated as the average expression ratio from the  
26  
27 77 different probes. Data for which at least 2 of the 3 probes gave a significant and non-  
28  
29 78 contradictory result were taken into account.

26  
27 79 All transcriptome data are MIAME compliant. The RAW data have been submitted to the  
28  
29 80 ArrayExpress database maintained at <http://www.ebi.ac.uk/microarray-as/ae/> under the  
30  
31 81 Accession number: E-MEXP-3236 for the rga experiment, E-MEXP-3237 for the rogB  
32  
33 82 experiment, and E-MEXP-3304 for the covSR experiment. It is worth mentioning that the  
34  
35 83 Rga microarray analysis revealed a number of genes that were strongly up-regulated in the  
36  
37 84  $\Delta rga$  mutant, such as *gbs0189-0190* (118- and 31-fold respectively), an observation that was  
38  
39 85 not confirmed by quantitative RT-PCR. We noticed that they belong to a set of about 140  
40  
41 86 genes for which a strong discrepancy was repeatedly observed for the WT strain between  
42  
43 87 independant microarrays experiments and/or qRT-PCR analysis. Our hypothesis is that the  
44  
45 88 expression of these genes is unexpectedly variable from one experiment to another one, even  
46  
47 89 though the growth conditions were cautiously monitored.

48  
49 90

#### 91 *Real-time quantitative PCR*

48  
49 92 Total RNAs extracted from NEM316, NEM316 $\Delta rga$ , and NEM316 $\Delta rogB$  were harvested at  
50  
51 93 an  $OD_{600nm} = 0.5$  as previously described<sup>28</sup>. cDNA synthesis was performed on DNase I  
52  
53 94 treated RNA (5 $\mu$ g) with random hexamers (Roche) using Superscript II RT (Invitrogen) as  
54  
55 95 recommended by the manufacturer. Real-time quantitative PCR was performed as described  
56  
57 96 previously<sup>28</sup> in a 25  $\mu$ l reaction volume containing cDNA, 12.5  $\mu$ l iQTM SYBR Supermix  
58  
59 97 (BioRad), and 1  $\mu$ l gene-specific primers (10  $\mu$ M each) (Table S2). Briefly, amplification,  
60  
98 98 detection, and analysis were performed with the MyiQ Single-Color Real-Time iCycler PCR

99 Detection System and the MyiQ Optical System Software (Bio-Rad). The specificity of the  
100 amplified products and the absence of primer dimers formation were verified by generating  
101 melting curves. The absence of contaminating genomic DNA was verified by testing each  
102 sample in control reactions without a prior reverse transcription step. The critical threshold  
103 cycle (CT) was defined for each sample. The expression levels of the tested genes were  
104 normalized using NEM316 *gyrA* gene whose transcription level did not vary under our  
105 experimental conditions. Each assay was performed in quadruplicate and repeated with two  
106 independent RNA samples. The change (n-fold) in the transcript level was calculated as  
107 described<sup>23</sup>.

#### 109 *DNase I footprinting assays*

110 DNA fragments corresponding to the promoter regions of *srr1* (409 bp amplified with primers  
111 *rga/srr1-F* and *-R*), *pilA* (295 bp, *rogB/pilA-F* and *-R*) were generated by PCR using Pwo  
112 polymerase (Roche) and the indicated oligonucleotide pairs (for sequence see Table S1).  
113 Radiolabeling of DNA fragments was performed as described previously<sup>5</sup>. Either purified  
114 Rga-6His or crude extracts (from *L. lactis/pAT18* and *L. lactis/pAT18-Rga*) binding to DNA  
115 ( $5 \times 10^4$  cpm per reaction) was performed at room temperature in a buffer containing 25 mM  
116  $\text{NaH}_2\text{PO}_4$  pH 8, 50 mM NaCl, 2 mM  $\text{MgCl}_2$ , 1 mM DTT and 10% glycerol. DNaseI treatment  
117 and migration of reactions mixtures were then performed as described previously<sup>5</sup>.

#### 119 *Primer extension reactions*

120 Total RNA was used as templates for primer extension reactions using radiolabeled specific  
121 primers of *rogB* (*gbs1479*), *pilA* (*gbs1478*), *rga* (*gbs1529*), and *srr1* (*gbs1530*) respectively  
122 (Table S1) as previously described<sup>7</sup>. The corresponding Sanger DNA sequencing reactions  
123 were carried out by using the same primers and PCR-amplified fragments containing the  
124 genes upstream regions with the Sequenase PCR product sequencing kit (USB, Cleveland,  
125 OH).

#### 127 *Immunoblotting analysis*

128 Cell wall protein extracts were prepared by harvesting 50 ml of bacteria in exponential phase  
129 ( $\text{OD}_{600}=0.5$ ). The bacterial pellet was washed in ice-cold phosphate buffer saline (PBS) and  
130 then in Tris-HCl buffer (50mM, pH 7.3) and finally the pellet was resuspended in the  
131 mutanolysin digestion buffer (Tris-HCl 50mM pH 7.3 supplemented with 20% sucrose and a

1  
2  
3 132 complete protease inhibitor cocktail (Roche)). Mutanolysin (Sigma) dissolved to 5,000 U ml<sup>-1</sup>  
4  
5 133 in potassium phosphate buffer (pH 6.2) was then added to the bacterial suspension to give a  
6  
7 134 final concentration of 100 U ml<sup>-1</sup> and samples were rotated for 2 h at 37°C. After centrifuging  
8  
9 135 at 13,000 g for 15 min at 4°C, supernatants corresponding to the cell wall extracts were  
10  
11 136 analyzed on SDS-PAGE or kept frozen at -20°C. For western blotting, proteins were boiled in  
12  
13 137 Laemmli sample buffer, resolved on Tris-Acetate Criterion XT gradient gels 4-12% SDS-  
14  
15 138 PAGE gels (BioRad) and transferred to nitrocellulose membrane (Hybond-C, Amersham).  
16  
17 139 For dot-blot analysis on whole bacteria, exponentially growing bacteria were washed in PBS  
18  
19 140 and resuspended in adjusted volumes of PBS to get similar OD<sub>600</sub> values. The bacteria were  
20  
21 141 loaded on nitrocellulose membrane, dried up for 20 min at room temperature, and then  
22  
23 142 blocked in PBS-milk 5% for 30 min. PilB and Srr1 were detected using specific rabbit  
24  
25 143 polyclonal antibodies obtained previously<sup>6,28</sup> at 1: 1000 dilution and as loading control *S.*  
26  
27 144 *agalactiae* was detected using a mouse antibody raised against formaldehyde killed BM110.  
28  
29 145 The secondary horseradish peroxidase (HRP)-coupled anti-rabbit secondary antibody  
30  
31 146 (Zymed) was used at 1:20000 dilution whereas the goat anti-mouse antibody was used at  
32  
33 147 1:10000 dilution. Detection was performed using the Western pico chemiluminescence kit  
34  
35 148 (Thermo Scientific). Image capture and analysis were done on GeneGnome imaging system  
36  
37 149 (Syngene).  
38  
39 150

#### 151 *Biofilm formation assays*

152 Bacterial attachment and surface growth on polystyrene microtiter plates were analyzed  
153 during growth of *S. agalactiae* in LB medium with 1% glucose supplemented with  
154 erythromycin when necessary. Overnight cultures grown in TH were used to inoculate LB  
155 glucose medium at 0.1 OD<sub>600</sub> and after a brief vortexing, 180 µl of cell suspension were  
156 dispensed into 96-wells plate (Costar 3799; Corning, Inc., NY) and incubated at 37°C for 2 h  
157 or 24 h. The OD<sub>600</sub> of each culture was measured to ensure that they have reached the  
158 stationary phase at similar cell densities, and the wells were washed twice in PBS using  
159 ELx50 washer (Biotek) and air-dried for 15 min. Biofilms were stained with 0.1% crystal  
160 violet for 30 min (100 µl per well) and the wells were washed twice with PBS and air-dried.  
161 The stained biomass was resuspended for quantification in ethanol/acetone (80:20) and A<sub>595</sub>  
162 was measured. The assay was performed in quadriplate.  
163

#### 164 *Cell culture and adherence assays*

1  
2  
3 165 The human cell lines A549 (ATCC CCL-185) from an alveolar epithelial carcinoma and TC7  
4 166 clone <sup>3</sup> established from the parental colon adenocarcinoma Caco-2, were cultured in  
5 167 Quantum 286 Medium (PAA). Cells were incubated in 10% CO<sub>2</sub> at 37° C and were seeded at  
6 168 a density of 2 to 5 x 10<sup>5</sup> cells per well in 24-well tissue culture plates. Monolayers were used  
7 169 after 24 - 72 h of incubation. Bacterial cultures from overnight cultures OD<sub>600</sub> of 2  
8 170 (approximately 6 10<sup>8</sup> CFU/ml) were washed once in PBS and resuspended in DMEM. Cells  
9 171 were infected at a multiplicity of infection (M.O.I) of 10 bacteria per cell for 1 h at 37 °C in  
10 172 10% CO<sub>2</sub>. The monolayers were then washed four times with PBS, and the cells were  
11 173 disrupted by the addition of 1 ml sterile deionized ice-cold water and repeated pipeting. Serial  
12 174 dilutions of the lysate were plated onto TH agar for count of viable bacteria. The percent of  
13 175 adherence was calculated as follows: (CFU on plate count / CFU in original inoculum) X 100.  
14 176 Assays were performed in triplicate and were repeated at least three times.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Acknowledgments

This work was supported in part by grant ANR-06-PATHO-001-01 from the Agence Nationale de la Recherche "ERA-NET PathoGenoMics" (to P. T.-C.) and by the Institut Pasteur. We thank Tarek Msadek for his help and constant support during the purification of recombinant Rga protein and Mathieu Brochet for the preliminary results using macroarrays. We also thank Agnès Fouet for critical reading of the manuscript.

For Peer Review

## References

1. **Brochet M., Couve E., Zouine M., Vallaeyts T., Rusniok C., Lamy M.C., Buchrieser C., Trieu-Cuot P., Kunst F., Poyart C. and others.** 2006. Genomic diversity and evolution within the species *Streptococcus agalactiae*. *Microbes Infect* 8: 1227-43.
2. **Campbell J.R., Hillier S.L., Krohn M.A., Ferrieri P., Zaleznik D.F., and Baker C.J.** 2000. Group B streptococcal colonization and serotype-specific immunity in pregnant women at delivery. *Obstet Gynecol* 96: 498-503.
3. **Chantret I., Rodolosse A., Barbat A., Dussaulx E., Brot-Laroche E., Zweibaum A., and Rousset M.** 1994. Differential expression of sucrase-isomaltase in clones isolated from early and late passages of the cell line Caco-2: evidence for glucose-dependent negative regulation. *J Cell Sci* 107 ( Pt 1): 213-25.
4. **Delmar P., Robin S., and Daudin J.J.** 2005. VarMixt: efficient variance modelling for the differential analysis of replicated gene expression data. *Bioinformatics* 21: 502-8.
5. **Derre I., Rapoport G., and Msadek T.** 1999. CtsR, a novel regulator of stress and heat shock response, controls clp and molecular chaperone gene expression in gram-positive bacteria. *Mol Microbiol* 31: 117-31.
6. **Dramsi S., Caliot E., Bonne I., Guadagnini S., Prevost M.C., Kojadinovic M., Lalioui L., Poyart C., and Trieu-Cuot P.** 2006. Assembly and role of pili in group B streptococci. *Mol Microbiol* 60: 1401-13.
7. **Dubrac S., Boneca I.G., Poupel O., and Msadek T.** 2007. New insights into the Walk/WalR (YycG/YycF) essential signal transduction pathway reveal a major role in controlling cell wall metabolism and biofilm formation in *Staphylococcus aureus*. *J Bacteriol* 189: 8257-69.
8. **Edwards M.S., and Baker C.J.** 2005. Group B streptococcal infections in elderly adults. *Clin Infect Dis* 41: 839-47.
9. **Edwards M.S., Nizet V., and Baker C.J.** 2006. Group B streptococcal infections. In: Remington JS, Klein JO, Wilson C, and Baker CJ, (eds.). *Infectious Diseases of the fetus and newborn infant*. Saunders Elsevier, pp. 403-464.
10. **Farley M.M.** 2001. Group B streptococcal disease in nonpregnant adults. *Clin Infect Dis* 33: 556-61.
11. **Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L., Msadek T., Zouine M., Couve E., Lalioui L., Poyart C. and others.** 2002. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol Microbiol* 45: 1499-513.
12. **Granok A.B., Parsonage D., Ross R.P., and Caparon M.G.** 2000. The RofA binding site in *Streptococcus pyogenes* is utilized in multiple transcriptional pathways. *J Bacteriol* 182: 1529-40.
13. **Gutekunst H., Eikmanns B.J., and Reinscheid D.J.** 2003. Analysis of RogB-controlled virulence mechanisms and gene repression in *Streptococcus agalactiae*. *Infect Immun* 71: 5056-64.
14. **Hondorp E.R., and McIver K.S.** 2007. The Mga virulence regulon: infection where the grass is greener. *Mol Microbiol* 66: 1056-65.
15. **Jiang S.M., Cieslewicz M.J., Kasper D.L., and Wessels M.R.** 2005. Regulation of virulence by a two-component system in group B streptococcus. *J Bacteriol* 187: 1105-13.

16. **Jones N., Bohnsack J.F., Takahashi S., Oliver K.A., Chan M.S., Kunst F., Glaser P., Rusniok C., Crook D.W., Harding R.M. and others.** 2003. Multilocus sequence typing system for group B streptococcus. *J Clin Microbiol* 41: 2530-6.
17. **Jones N., Oliver K.A., Barry J., Harding R.M., Bisharat N., Spratt B.G., Peto T., and Crook D.W.** 2006. Enhanced invasiveness of bovine-derived neonatal sequence type 17 group B streptococcus is independent of capsular serotype. *Clin Infect Dis* 42: 915-24.
18. **Konto-Ghiorgi Y., Mairey E., Mallet A., Dumenil G., Caliot E., Trieu-Cuot P., and Dramsi S.** 2009. Dual role for pilus in adherence to epithelial cells and biofilm formation in *Streptococcus agalactiae*. *PLoS Pathog* 5: e1000422.
19. **Kreikemeyer B., McIver K.S., and Podbielski A.** 2003. Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. *Trends Microbiol* 11: 224-32.
20. **Kuipers O.P., de Ruyter P.G.G.A., Kleerebezem M., and de Vos W.M.** 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. *Journal of Biotechnology* 64: 15-21.
21. **Lamy M.C., Zouine M., Fert J., Vergassola M., Couve E., Pellegrini E., Glaser P., Kunst F., Msadek T., Trieu-Cuot P. and others.** 2004. CovS/CovR of group B streptococcus: a two-component global regulatory system involved in virulence. *Mol Microbiol* 54: 1250-68.
22. **Lin W.J., Walthers D., Connelly J.E., Burnside K., Jewell K.A., Kenney L.J., and Rajagopal L.** 2009. Threonine phosphorylation prevents promoter DNA binding of the Group B Streptococcus response regulator CovR. *Mol Microbiol* 71: 1477-95.
23. **Livak K.J., and Schmittgen T.D.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402-8.
24. **Marquis R.E., Bender G.R., Murray D.R., and Wong A.** 1987. Arginine deiminase system and bacterial adaptation to acid environments. *Appl Environ Microbiol* 53: 198-200.
25. **McIver K.S.** 2009. Stand-alone response regulators controlling global virulence networks in *Streptococcus pyogenes*. *Contrib Microbiol* 16: 103-19.
26. **McIver K.S., Heath A.S., Green B.D., and Scott J.R.** 1995. Specific binding of the activator Mga to promoter sequences of the *emm* and *scpA* genes in the group A streptococcus. *J Bacteriol* 177: 6619-24.
27. **McIver K.S., and Myles R.L.** 2002. Two DNA-binding domains of Mga are required for virulence gene activation in the group A streptococcus. *Mol Microbiol* 43: 1591-601.
28. **Mistou M.Y., Dramsi S., Brega S., Poyart C., and Trieu-Cuot P.** 2009. Molecular dissection of the *secA2* locus of group B Streptococcus reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence. *J Bacteriol* 191: 4195-206.
29. **Musser J.M., Mattingly S.J., Quentin R., Goudeau A., and Selander R.K.** 1989. Identification of a high-virulence clone of type III Streptococcus agalactiae (group B Streptococcus) causing invasive neonatal disease. *Proc Natl Acad Sci U S A* 86: 4731-5.
30. **Podbielski A., Woischnik M., Leonard B.A., and Schmidt K.H.** 1999. Characterization of *nra*, a global negative regulator gene in group A streptococci. *Mol Microbiol* 31: 1051-64.
31. **Que Y.A., Haefliger J.A., Francioli P., and Moreillon P.** 2000. Expression of *Staphylococcus aureus* clumping factor A in *Lactococcus lactis* subsp. *cremoris* using a new shuttle vector. *Infect Immun* 68: 3516-22.



- 1  
2  
3 32. **Rajagopal L.** 2009. Understanding the regulation of Group B Streptococcal virulence  
4 factors. *Future Microbiol* 4: 201-21.  
5 33. **Reiner A., Yekutieli D., and Benjamini Y.** 2003. Identifying differentially expressed  
6 genes using false discovery rate controlling procedures. *Bioinformatics* 19: 368-75.  
7 34. **Rinaudo C.D., Rosini R., Galeotti C.L., Berti F., Necchi F., Reguzzi V., Ghezze  
8 C., Telford J.L., Grandi G., and Maione D.** 2010. Specific involvement of pilus  
9 type 2a in biofilm formation in group B Streptococcus. *PLoS One* 5: e9216.  
10 35. **Rosini R., Rinaudo C.D., Soriani M., Lauer P., Mora M., Maione D., Taddei A.,  
11 Santi I., Ghezze C., Brettoni C. and others.** 2006. Identification of novel genomic  
12 islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*. *Mol*  
13 *Microbiol* 61: 126-41.  
14 36. **Sambrook J., Fritsch E.F., and Maniatis T.** 1989. *Molecular Cloning: a Laboratory*  
15 *Manual*. Cold Spring Harbor Laboratory, NY, New-York.  
16 37. **Samen U., Heinz B., Boisvert H., Eikmanns B.J., Reinscheid D.J., and Borges F.**  
17 2011. Rga is a regulator of adherence and pili formation in *Streptococcus agalactiae*.  
18 *Microbiology*.  
19 38. **Schuchat A.** 1998. Epidemiology of group B streptococcal disease in the United  
20 States: shifting paradigms. *Clin Microbiol Rev* 11: 497-513.  
21 39. **Tazi A., Disson O., Bellais S., Bouaboud A., Dmytruk N., Dramsi S., Mistou  
22 M.Y., Khun H., Mechler C., Tardieux I. and others.** 2010. The surface protein  
23 HvgA mediates group B streptococcus hypervirulence and meningeal tropism in  
24 neonates. *J Exp Med* 207: 2313-22.  
25 40. **Tettelin H., Masignani V., Cieslewicz M.J., Donati C., Medini D., Ward N.L.,  
26 Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S. and others.** 2005. Genome  
27 analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for  
28 the microbial "pan-genome". *Proc Natl Acad Sci U S A* 102: 13950-5.  
29 41. **Yang Y.H., Dudoit S., Luu P., Lin D.M., Peng V., Ngai J., and Speed T.P.** 2002.  
30 Normalization for cDNA microarray data: a robust composite method addressing  
31 single and multiple slide systematic variation. *Nucleic Acids Res* 30: e15.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Figure Legends

**Fig. 1.** Comparison of the Rga regulon to the CovSR and RogB regulons using Cluster 3.0 and Java Tree View representation. Among the 97 differentially regulated genes in the Rga regulon, a subset of 61 genes whose expression varies in the CovSR or RogB microarrays are shown. Up-regulated genes are indicated in red, down-regulated genes are indicated in green, and non-variable genes are shown in black.

**Fig. 2.** Expression of the genes *pilB* and *srr1* in *S. agalactiae*  $\Delta$ *rogB* (top panel) and  $\Delta$ *rga* (bottom panel). The transcriptional analysis was performed by quantitative RT-PCR and normalized relative to *gyrA* gene. The figure indicates the calculated ratio of the expression in the mutant strains relative to that in the NEM316 WT strain.

**Fig. 3.** Analysis of surface exposure of Srr1 and PilB proteins in WT NEM316,  $\Delta$ *rga* and  $\Delta$ *rogB* strains by immunodot-blot. 2  $\mu$ l and 6  $\mu$ l of whole bacterial cells suspension in PBS (OD<sub>600nm</sub>=1) harvested in exponential phase (OD<sub>600nm</sub>=0.5) were spotted on nitrocellulose. Membranes were hybridized with specific anti-Srr1 and anti-PilB rabbit polyclonal antibodies. Anti-GBS antiserum raised in mice was used for loading control. Membrane were next incubated with HRP-conjugated secondary antibodies and revealed using the Western pico chemiluminescence kit.

**Fig. 4.** Western blot analysis of cell wall anchored pili in three *S. agalactiae* strains overexpressing *rogB* or *rga*. NEM316 and its isogenic  $\Delta$ *rgalrogB* mutant strains (panel A) and *S. agalactiae* 2603V/R strain (panel B) harboring the empty vector or the plasmids carrying *rogB* or *rga*. All the strains were grown in the presence of erythromycin. Mutanolysin extracted cell wall associated proteins were separated on 4-12% gradient SDS-PAGE and detected after transfer on nitrocellulose with specific anti-PilB (panel A) or anti-Sag1407 (panel B) antisera. The molecular weight marker are indicated on the left (kDa).

**Fig. 5.** DNase I footprinting analysis of Rga binding to the *srr1* promoter region. A : DNase I protection assay on the *srr1* promoter region using crude extracts of *Lactococcus lactis* harboring either the pAT18 empty vector or the pAT18-*rga* plasmid. Lanes contain approximately 0.5 pmol ( $5 \times 10^4$  cpm per reaction) of labeled template strand of *srr1* (-176 to +232 relative to the transcription start site). Lane 1: no protein; lanes 2 and 5: 25  $\mu$ g; lanes 3

1  
2  
3 and 6: 50  $\mu\text{g}$  ; lanes 4 and 7 : 100  $\mu\text{g}$  of *L. lactis*/pAT18 crude extract in lanes 2, 3 and 4 and  
4  
5 *L. lactis*/pAT18-*rga* crude extract in lanes 5, 6 and 7. Lane 8: A+G Maxam and Gilbert  
6  
7 reaction. Brackets indicate regions protected by Rga from DNase I cleavage. B: Nucleotide  
8  
9 sequences of the *rga* promoter region. Initiation site of transcription has been mapped (Fig.  
10  
11 S2) and indicated as +1, and the -10 and -35 regions are shaded. Regions protected from  
12  
13 DNase I digestion upon Rga binding are indicated.

14  
15 **Fig. 6. Role of transcriptional regulators Rga and RogB in biofilm formation.** *S.*  
16 *agalactiae* NEM316 and its isogenic  $\Delta\text{rga}$  and  $\Delta\text{rogB}$  mutants (panel A),  $\Delta\text{rogB}$  or  $\Delta\text{rga}$   
17 complemented strains (panel B), NEM316, NEM316 $\Delta\text{rga}/\text{rogB}$  and 2603V/R overexpressing  
18 *rogB* or *rga* as compared to the corresponding strain harboring the empty plasmid  
19 (vec). Bacteria were grown in 96 wells polystyrene plates in LB supplemented with 1%  
20 glucose at 37°C for 24h. Adherent bacteria were stained with crystal violet (CV) and  
21 quantification was performed by measuring absorbance at 595 nm. For bacteria harboring the  
22 plasmids, the experiments of biofilm formation were performed in the presence of  
23 erythromycin which affects the final biomass and results in overall less biofilm (compare  
24 panel B and A). Results displayed in the histograms correspond to the average of three  
25 independent experiments (error bars are standard deviations).

26  
27  
28  
29  
30  
31  
32  
33  
34 **Fig. 7. Adhesion of *S. agalactiae* NEM316 and its isogenic  $\Delta\text{rogB}$ ,  $\Delta\text{rga}$ , and  $\Delta\text{rga}/\text{rogB}$**   
35 **mutants to the alveolar epithelial A549 cell line.** Monolayers were infected at a multiplicity  
36 of infection of 10 bacteria per cell for 1 hour at 37°C. The adherence frequencies were  
37 calculated from the numbers of bacteria remained attached to the cell after the incubation  
38 period and three washes with respect to the number of inoculated bacteria. The level of  
39 adherence of the WT strain is arbitrarily reported as 100 and the level of adherence of the  
40 various mutants are relative values. The results are presented as mean value ( $\pm\text{SD}$ ) from 3  
41 independent experiments, each strain tested in triplicate.

42  
43  
44  
45  
46  
47  
48  
49 **Fig. 8. Working model for Rga regulation and its integration in other regulatory**  
50 **networks.** Strong transcriptional activation of *srr1* and PI-2a operon by Rga (black thick  
51 lines) and low activation of the PI-2a operon by RogB (grey thin lines). RogB also activates  
52 its own transcription and slightly represses *srr1*. At an upper level, the CovS/R two-  
53 component system, central for virulence in GBS, negatively regulates *rga* transcription.  
54  
55  
56  
57  
58  
59  
60

**Supplementary Figures:**

**Fig. S1. Comparison of RogB (A) and Rga (B) deduced amino acid sequences in the sequenced GBS strains.** NEM316, 2603V/R, A909, H36B, 515, CJB111 and 18RS21.

**Fig. S2. Determination of transcriptional start sites of the *pilA*, *rogB*, *srr1* and *rga* genes by primer extension analysis (A).** Total RNA was extracted from the wild type NEM316 strain and primers specific for each gene (see Table S1) were radiolabelled and used for primer extensions. Sanger sequencing reactions (GATC) were carried out on PCR fragments corresponding to each gene. Nucleotide sequences corresponding to the upstream regions of each gene are represented. The initiation codons are underlined. Initiation sites of transcription are indicated as +1, and the -10 and -35 regions are shaded. Two transcripts have been detected for the *rogB* gene, one beginning upstream the gene and thought to start at the initiation site of transcription, and one beginning downstream the initiation site of translation (indicated by an asterisk) and thought to be a degradation product. B. Genetic organization of the *srr1-rga*, and PI-2A-*rogB* loci. Distances between initiation starts of transcription and translation are indicated in base pairs. The entire DNA sequence of the *srr1-rga* promoter region is shown with the transcription start sites indicated by an arrow (+1) and in bold the corresponding -10 and -35 regions. Regions protected from DNase I digestion upon Rga binding are indicated in shaded grey.

**Figure S3:** Expression of the GBS NEM316 PI-2A pilus operon in *Lactococcus lactis* strain NZ9000. A- Schematic representation of the PI-2a pilus operon expressed under the lactococcal promoter p23 in the plasmid pOri23. B- Western blotting of cell-wall anchored proteins of recombinant *L. lactis* strains expressing the three structural pilin subunits (*pilA*, *pilB*, or *pilC*) or the entire operon PI-2a. PilB, the major pilin subunit is detected by immunoblotting with specific anti-PilB (Gbs1477). PilB monomer is indicated by a black arrow (m) and polymers are indicated by a line with p. C- Immunogold scanning micrographs using the specific anti-PilB antiserum showing the formation of pili in the recombinant *L. lactis* expressing the entire PI-2a operon. D- Biofilm formation in LB glucose 1% at 24 h and 30°C. E- Adherence of the recombinant *L. lactis* strains to intestinal epithelial cells TC7.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

**Table 1.** List of the annotated genes most differentially expressed in the  $\Delta rga$ ,  $\Delta rogB$ ,  $\Delta CovRS$  mutants relative to the parental strain NEM316<sup>a</sup>.

Gene <sup>b</sup>	Functional classification <sup>c</sup>	Alternative Name, Function	Ratio $\Delta rga/WT$ <sup>d</sup>	Ratio $\Delta rogB/WT$	Ratio $\Delta covRS/WT$ <sup>d</sup>
<i>Genes activated by Rga</i>					
<b>gbs1529*</b>	1.8	<i>srr1</i> , Ser rich repeat glycoprotein (LPXTG)	0.04	3.09	4.97 <sup>#</sup>
<b>gbs1474*</b>	1.8	<i>pilC</i> , pilus-anchoring subunit (LPXTG)	0.09	0.98	3.91 <sup>#</sup>
<b>gbs1475*</b>	1.8	<i>srtC4</i> , pilus-associated class C sortase	0.11	0.69	3.88 <sup>#</sup>
<b>gbs1476*</b>	1.8	<i>srtC3</i> , pilus-associated class C sortase	0.09	0.69	3.57 <sup>#</sup>
<b>gbs1477*</b>	1.8	<i>pilB</i> , major pilin subunit (LPXTG)	0.03 <sup>&amp;</sup>	0.4	3.81 <sup>#</sup>
<b>gbs1478*</b>	1.8	<i>pilA</i> , pilus-associated adhesin (LPXTG)	0.07 <sup>&amp;</sup>	0.4	3.36 <sup>#</sup>
<b>gbs1479</b>	3.5.2	<i>rogB</i> , transcriptional regulator RALP family	0.31	0.3	0.88
<b>gbs1432</b>	2.3	Carbamoyl-phosphate syntase, LU	0.37	0.88	1.01
<b>gbs1433</b>	2.3	Carbamoyl-phosphate syntase, SU	0.27	0.79	1.02
<b>gbs1434</b>	3.5.2	Pyrimidine operon regulatory protein	0.30	0.75	0.99
<b>gbs1435</b>	3.6	Ribosomal pseudouridine synthase, RluA	0.08	0.89	1.12
<b>gbs1436*</b>	1.6	<i>Isp</i> , Signal peptidase II	0.06	0.78	1.22
<b>gbs1437*</b>	3.5.2	Transcriptional regulator, LysR family	0.07	0.73	1.11
<b>gbs2083</b>	1.2	Arginine/ornithine antiporter	0.07	0.94	0.12
<b>gbs2084*</b>	2.2	Carbamate kinase (CK)	0.09	1	0.11
<b>gbs2085*</b>	2.2	Ornithine carbamoyltransferase 1	0.11	0.94	0.22
<b>gbs2122</b>	2.2	Arginine deiminase	0.44	0.61	0.23
<b>gbs2123</b>	5.2	Acetyl transferase GNAT family	0.45	0.27	0.17

<b>gbs2124</b>	2.2	Ornithine carbamoyltransferase 2	0.18	0.31	0.15
<b>gbs2125</b>	1.2	Arginine/ornithine antiporter	0.21	0.35	0.34
<b>gbs2126</b>	2.2	Carbamate kinase (ArcC)	0.26	0.35	0.57
<b>gbs1995</b>	5.2	Phospholipid phosphatase	0.2 <sup>&amp;</sup>	0.36	0.78
<b>gbs0322</b>	2.2	<i>cysK</i> , Cysteine synthase	0.3	2.06	0.88
<b>gbs1859</b>	2.3	PurA, similar to adenylosuccinate synthase	0.33	1.14	0.39
<b>gbs1376</b>	4.1	<i>clpL</i> , ATP dependant Clp protease	0.37 <sup>&amp;</sup>	0.28	0.69
<b>gbs1399</b>	5.2	<i>reP</i> , RelA/SpoT domain protein	0.47 <sup>&amp;</sup>	0.33	0.79
<b>gbs1734</b>	1.2	<i>scrA</i> , PTS system, sucrose-specific EIIBC	0.24	0.29	0.45
<b>gbs1330</b>	1.2	<i>lacE</i> , lactose-specific PTS system, EIIBC	0.34	0.63	0.35
<b>gbs1331</b>	1.2	Lactose-specific PTS system, EIIA	0.45	0.59	0.68
<b>gbs1444</b>	5.2	Ser-rich repeat protein	0.35	0.67	0.46
<b>gbs2018</b>	1.8	BibA, adhesin (LPXTG)	0.49	0.66	77.17 <sup>#</sup>
<b>gbs1605</b>	5.2	Glycosyltransferase, family 2	0.44 <sup>&amp;</sup>	0.53	101.19 <sup>#</sup>
<b>gbs1606</b>	5.2	Glycosyltransferase, family 2	0.43 <sup>&amp;</sup>	0.48	77.63 <sup>#</sup>
<b>gbs1607</b>	5.2	Spore coat protein Coth	0.31 <sup>&amp;</sup>	0.55	68.83 <sup>#</sup>
<b>gbs1608</b>	6	Membrane-associated protein	0.57	0.5	72.43 <sup>#</sup>
<b>gbs1609</b>	6	VTC-domain containing protein	0.46	0.57	76.51 <sup>#</sup>

<sup>a</sup> RNAs were extracted from bacteria collected at mid-exponential phase of growth in TH broth (OD<sub>600</sub> = 0.3).

<sup>b</sup> Stars (\*) indicate the few chosen genes whose differential expression has been validated by qRT-PCR, thereby confirming the microarray results.

<sup>c</sup> The functional classification is that of the SagaList server: <http://genolist.pasteur.fr/SagaList/help/function-codes.html>.

<sup>d</sup> #, Rga-dependent CovRS regulation; &, RogB-dependent Rga regulation.

For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

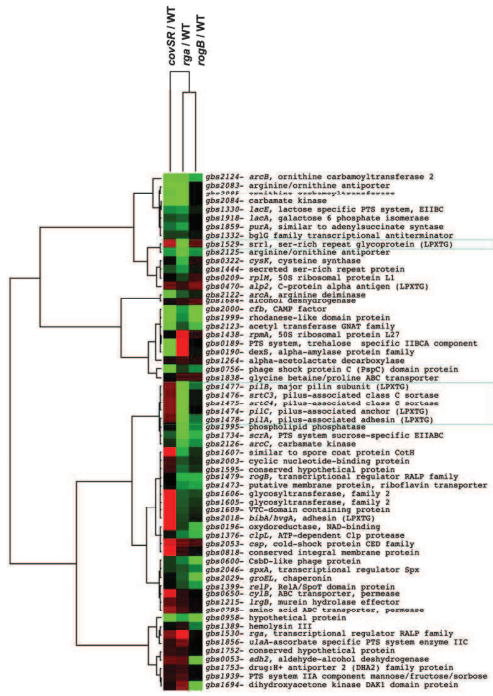


Figure 1

170x127mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

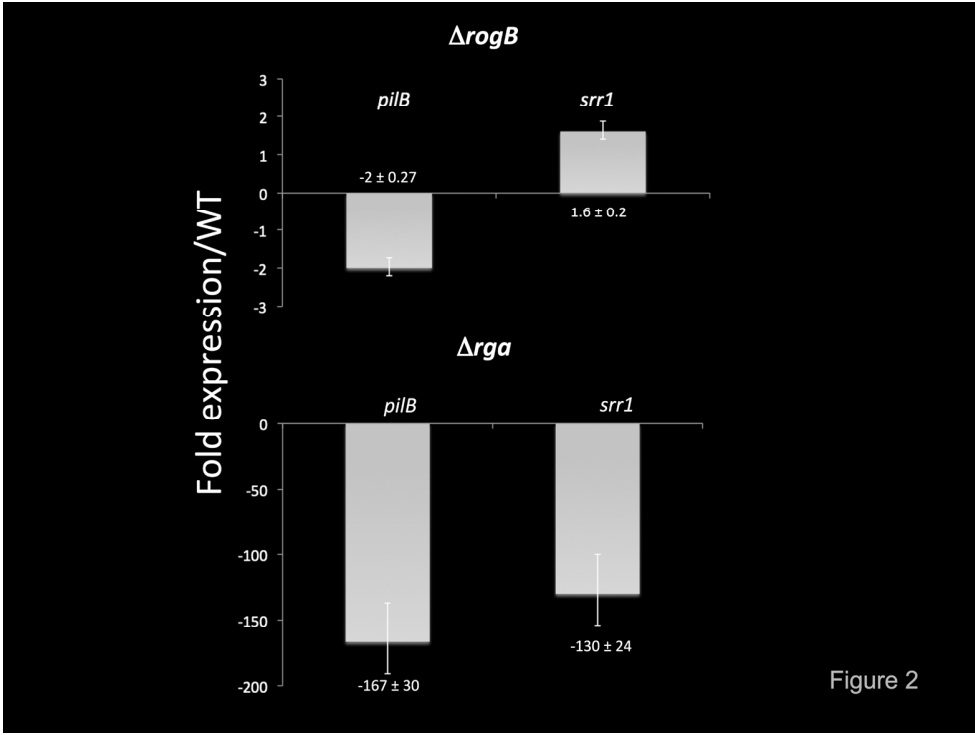


Figure 2

199x150mm (300 x 300 DPI)

Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

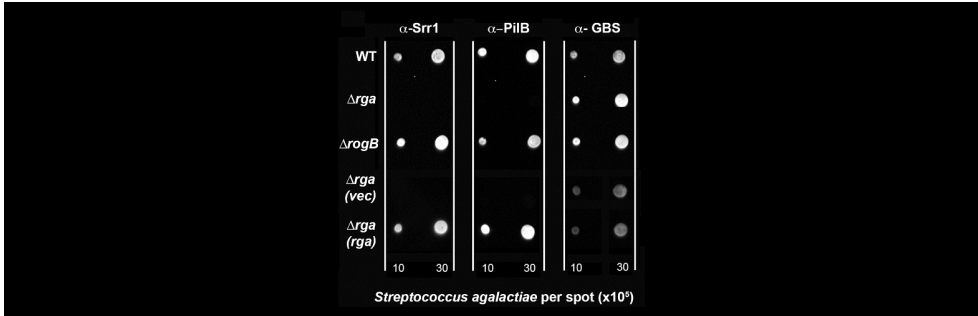


Figure 3

350x262mm (300 x 300 DPI)

Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

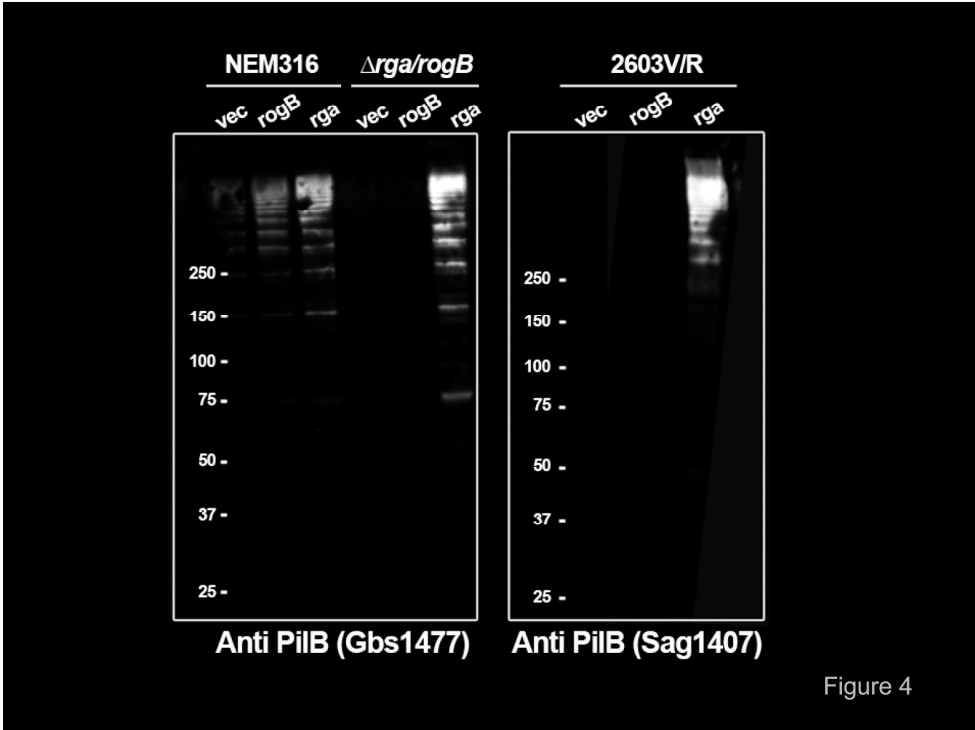
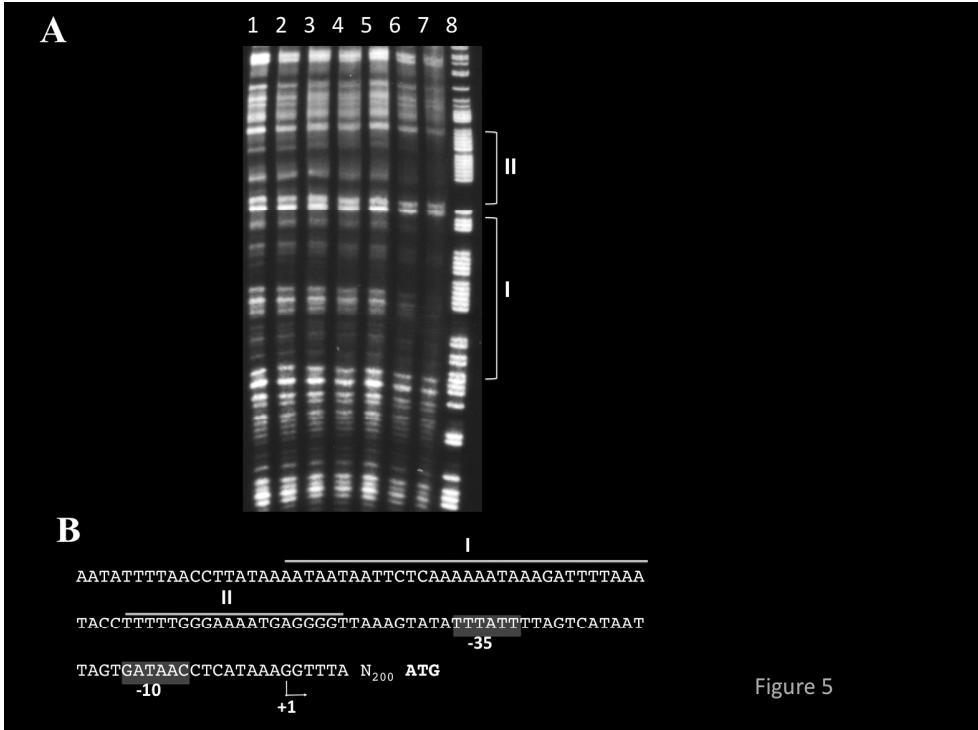


Figure 4

350x262mm (300 x 300 DPI)

Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



350x262mm (300 x 300 DPI)

Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

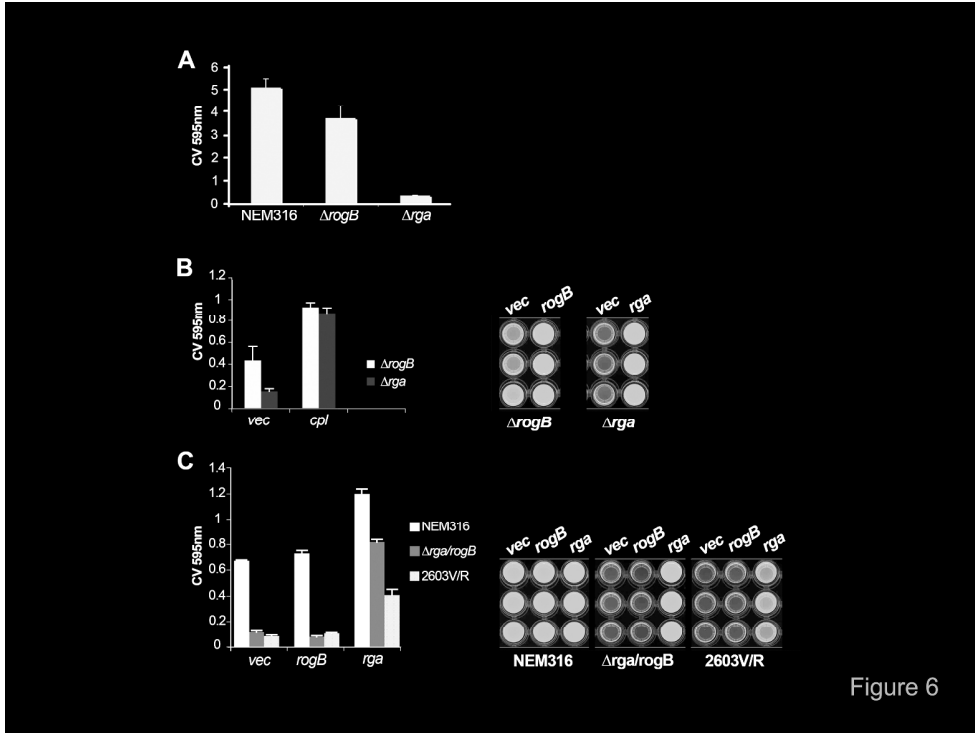
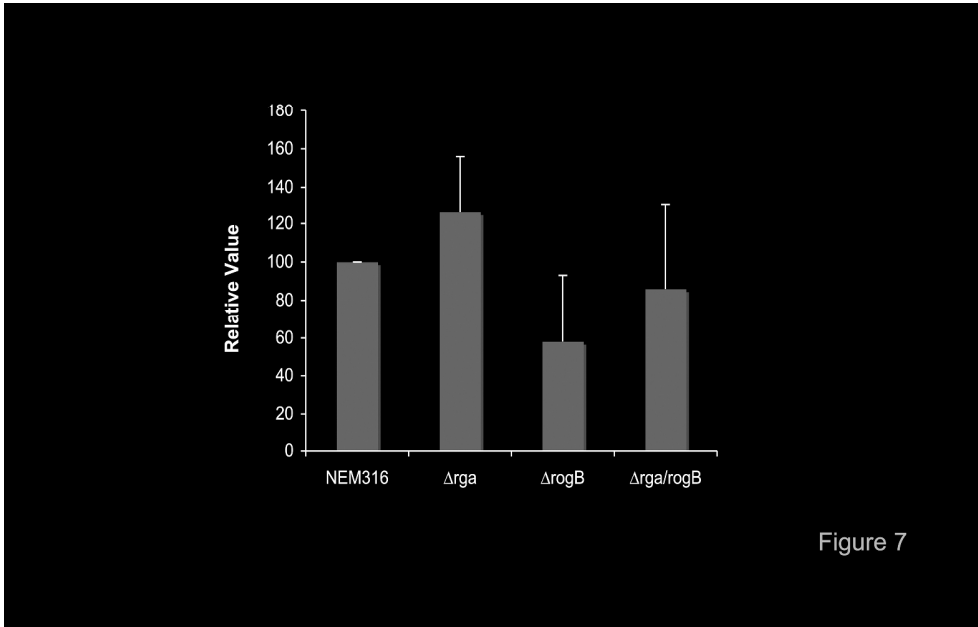


Figure 6

350x262mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



299x224mm (300 x 300 DPI)

Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

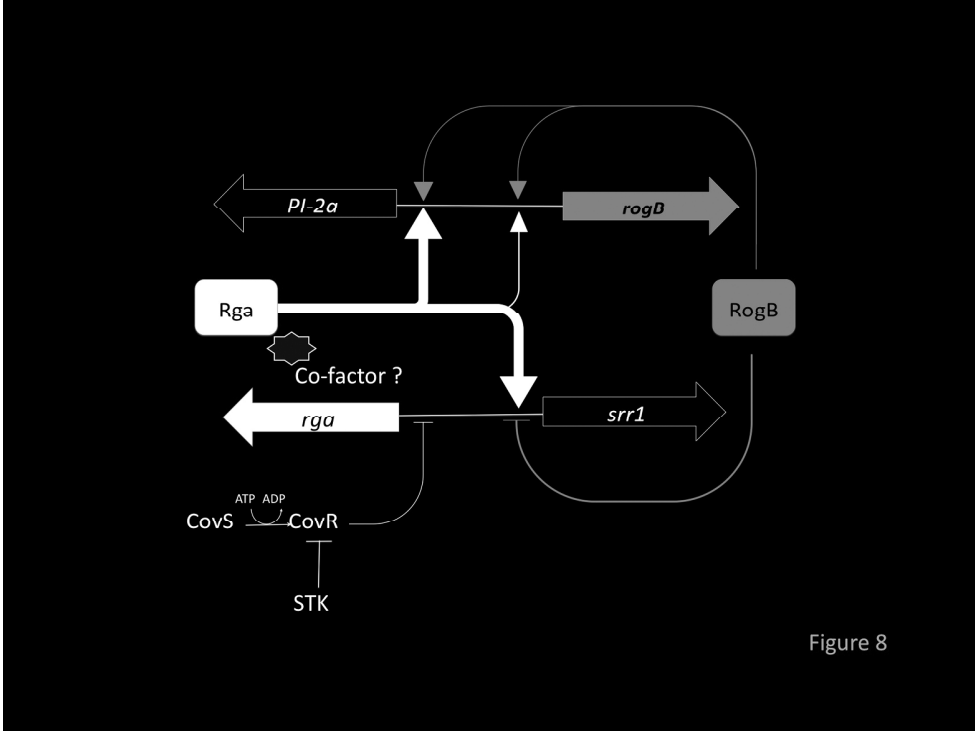


Figure 8

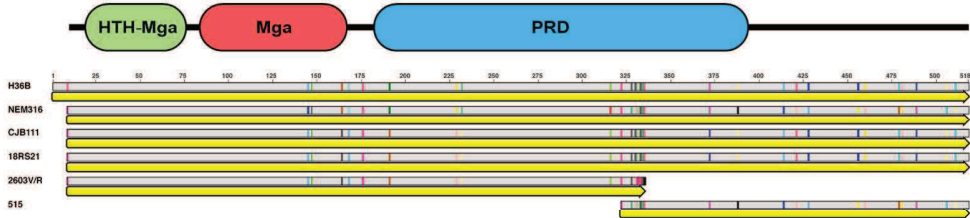
350x262mm (300 x 300 DPI)

Review

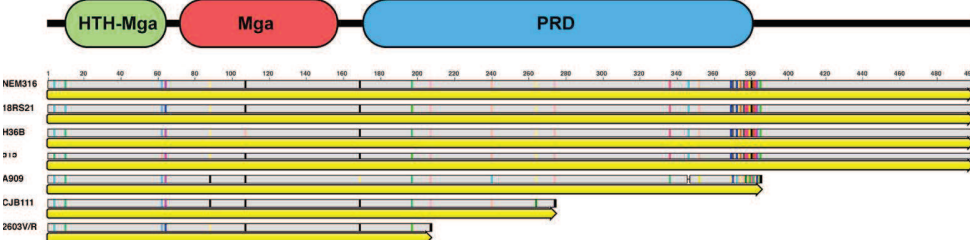


1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

A) RogB



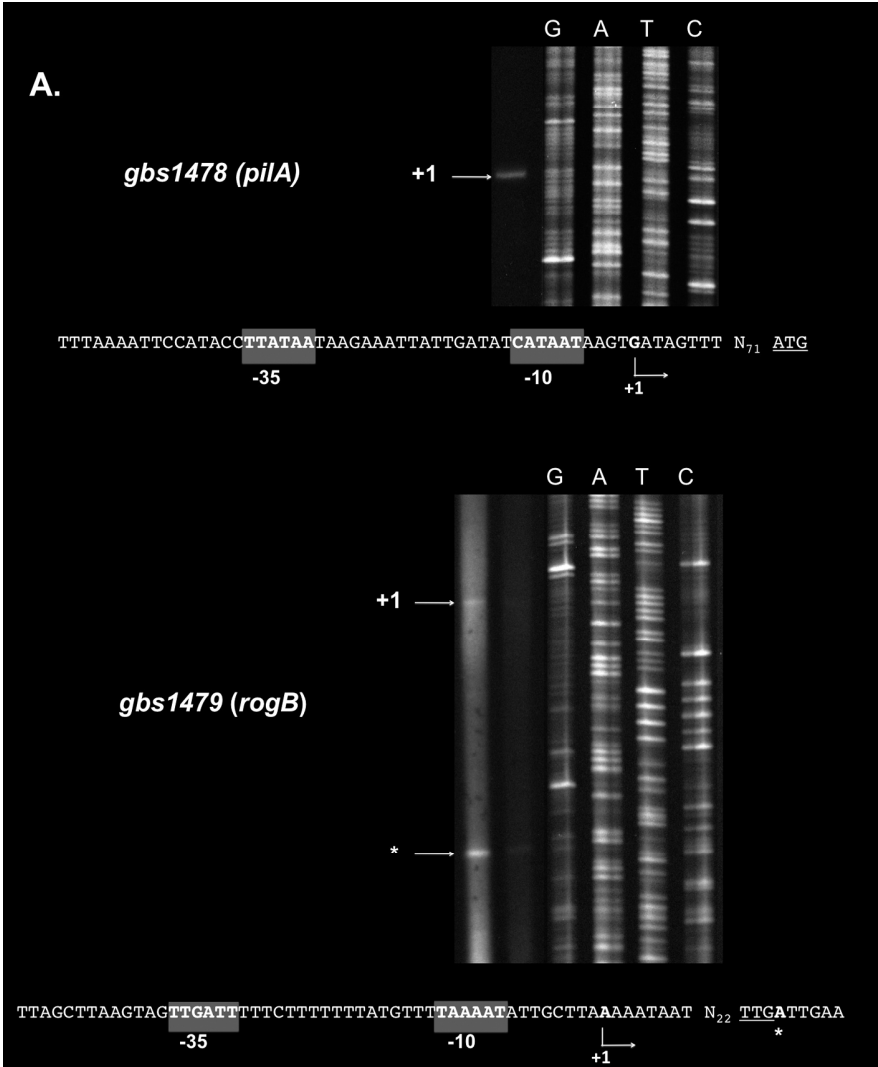
B) Rga



189x122mm (300 x 300 DPI)

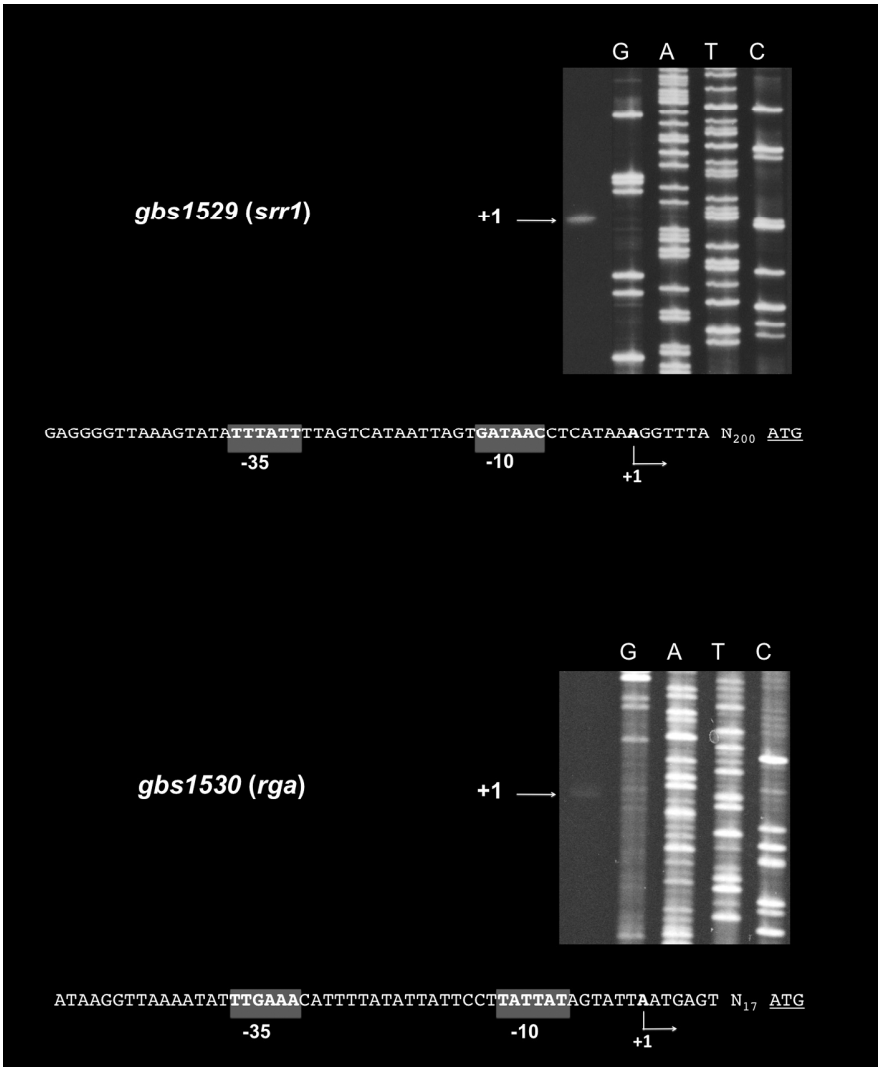
Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



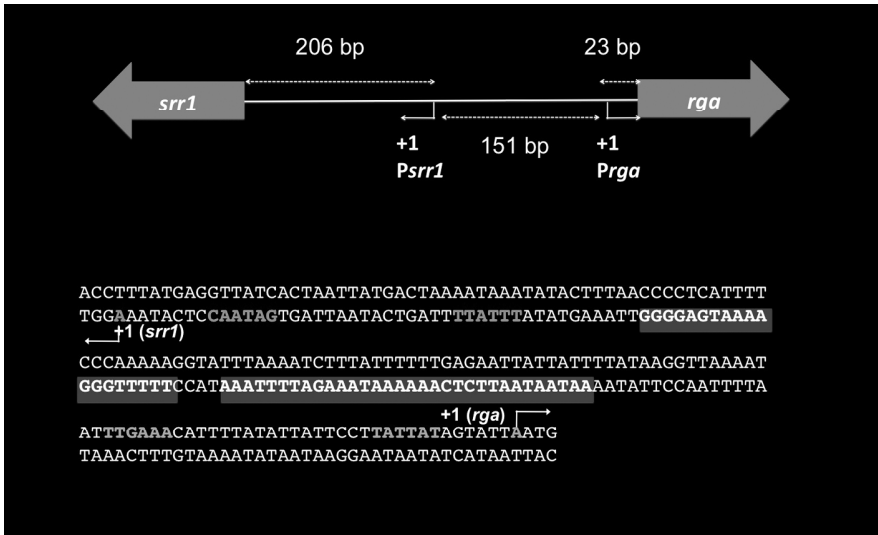
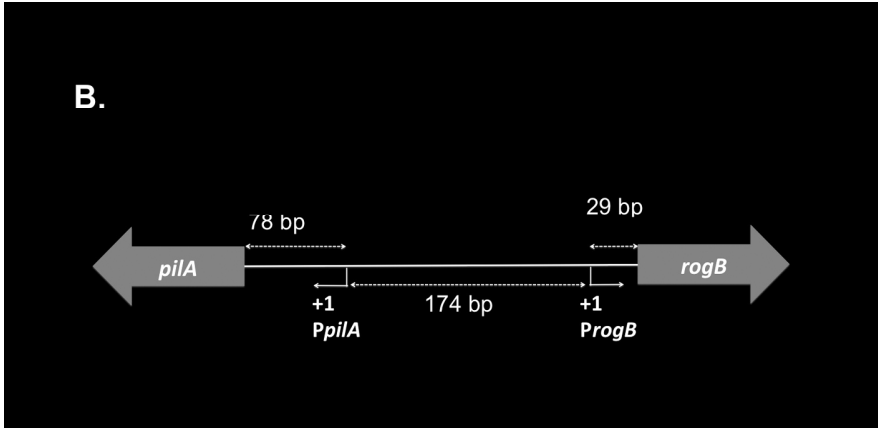
262x350mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



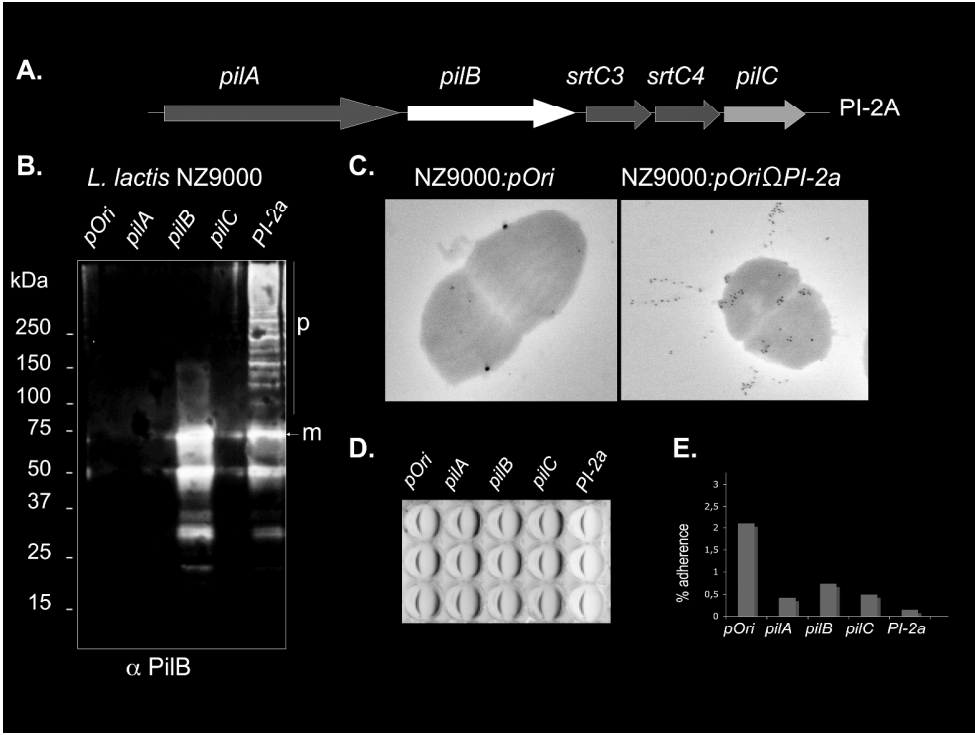
262x350mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



262x350mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



350x262mm (300 x 300 DPI)

Review

### (c) Résultats complémentaires concernant le troisième régulateur RofA-like Gbs1426.

L'ensemble des résultats précédents met en évidence les interrelations entre les différents réseaux de régulation contrôlant la biosynthèse des composants de la paroi. L'analyse du génome de la souche NEM316 indique l'existence d'un troisième gène codant pour un régulateur de la famille RofA, *gbs1426*. L'analyse de la distribution des trois gènes codant pour des régulateurs de type RofA au sein des différents isolats de *S. agalactiae* montre que seul *gbs1426* fait partie du core genome. Les souches ST17, par exemple, expriment des allèles différents du gène *srr* (*srr2*) et du pilus 2 (PI-2b), qui ne sont pas associés à des gènes régulateurs. Ces différences mettent en évidence la diversité des régulations au sein de l'espèce. La délétion de *gbs1426* modifie l'expression de 172 gènes, dont 10 gènes codant pour des protéines ribosomales, le gène codant pour l'ADN polymérase *polA*, la protéine de division cellulaire FtsZ, des gènes codant pour des sous unités de l'ATPase FOF1 (*atpB*, *atpF* et *atpH*), le gène *gbs0911* codant une des protéine du locus CRISPR. De façon surprenante, l'expression des gènes *pilA*, *pilB* et des deux sortases est également induite dans le mutant *gbs1426* sans que l'on constate de modification d'expression pour *rga* et *rogB*, suggérant un effet opposé de ce régulateur sur l'expression de cet opéron. L'expression de l'opéron pilus est également fortement induite dans le mutant du système à deux composants GraRS et le gène *pilA* est retrouvé différentiellement exprimé dans plus de 50% des expériences de transcriptome réalisées au cours de ma thèse. Ces observations suggèrent que l'expression du pilus PI-2a est probablement très sensible à des altérations des conditions de croissance.

## 3. Analyse combinée de l'ensemble des données de transcriptome

### (1) Contexte des différents projets réalisés en collaborations

Dans le cadre de différentes collaborations, j'ai réalisé l'analyse de 34 comparaisons transcriptomiques (Tableau 8). L'ensemble de ces analyses peut être regroupé en six projets majeurs:

#### a) Analyse de la réponse à l'absence de D-alanylation

Dans le cadre de ce premier projet, j'ai étudié la réponse de *S. agalactiae* à l'absence de

D-alanylation et à la présence de peptides antimicrobiens et j'ai identifié les systèmes de régulation potentiellement impliqués dans la régulation de l'opéron *dlt*. Pour cela, outre les 7 comparaisons indiquées dans l'article 1, j'ai également analysé l'effet de la double mutation  $\Delta dltR \Delta ciaRH$  et l'effet de l'ajout de colistine qui n'ont pas été inclus dans l'article. Finalement, j'ai aussi analysé la réponse de *S. agalactiae* à la délétion d'un autre système à deux composants, *gbs0963-0964*, un homologue de GraRS qui, chez *S. aureus*, est impliqué dans la régulation de la transcription de l'opéron *dlt* [342]. Cette analyse montre que l'inactivation de ce régulateur ne semble pas avoir d'effet sur l'expression de l'opéron *dlt* chez *S. agalactiae* dans les conditions testées ni sur le phénotype de une sensibilité à la colistine.

### b) **L'analyse des régulateurs de la famille RofA**

Ce projet, réalisé en collaboration avec Shaynoor Dramsi et Michel-Yves Mistou, consiste en l'analyse des trois régulateurs de la famille RofA Rga, RogB et Gbs1426 rapportée précédemment.

### c) **Rôle de AdtA sur CovRS**

Ce projet, réalisé en collaboration avec Arnaud Firon (Institut Pasteur) porte sur l'analyse du rôle de la protéine AdtA (pour Abi-Domaine Transmembrane protein A) au sein du réseau de régulation CovRS. La protéine AdtA présente 8 domaines transmembranaires et un domaine Abi (Pfam02517). Ce domaine Abi est présent dans de nombreuses protéines bactériennes, chaque génome bactérien pouvant coder pour plusieurs de ces protéines (jusqu'à 15 dans le cas de *Bacillus anthracis* [343]) pouvant. Cependant leur fonction, chez les procaryotes, n'a pas encore été caractérisée. Chez *S. agalactiae* les gènes codent pour des protéines avec des domaines Abi, sont au nombre de cinq dont *gbs1037*, et *gbs1532 (adtA)*. Deux raisons ont conduit à l'analyse du rôle de AdtA dans le réseau de régulation CovRS. Premièrement, le mutant *adtA* présente une diminution de pigmentation sur milieu grenada, ainsi qu'une faible activité hémolytique résultant probablement d'une diminution d'expression de l'opéron *cyl*, dont l'expression est régulée par CovR. Deuxièmement, l'expression du gène *adtA* est induite dans le mutant  $\Delta covRS$ .

d) ***S. agalactiae* en culture aérée versus statique.**

*S. agalactiae* est une bactérie aéro-anaérobie facultative, elle est à la fois capable de fermentation et de respiration aérobie, si le milieu de culture contient une quinolone, et de l'hème que *S. agalactiae* n'a pas la capacité de synthétiser [344, 345]. L'inactivation du gène *cydA*, codant pour une sous-unité de la cytochrome oxydase terminale, diminue fortement la croissance dans le sang et la virulence de *S. agalactiae* dans des modèles néonataux de septicémie chez le rat [344], révélant l'importance de la respiration dans la dissémination de *S. agalactiae*. Un autre gène, *nox2* (*gbs0946*), codant pour la NADH oxydase, est essentiel pour la croissance aérobie et la résistance au stress oxydatif de *S. agalactiae* en milieu de culture conventionnel. La NADH oxydase est une enzyme qui catalyse la réduction de l'O<sub>2</sub> en H<sub>2</sub>O en utilisant le NADH comme donneur d'électron [346]. Ce projet réalisé en collaboration avec Philippe Gaudu a porté sur l'analyse, dans des conditions de culture aérée ou statique, de l'impact sur l'expression génique de la délétion du régulateur transcriptionnel Rex (pour "redox sensing") dont l'activité dépend du ratio NADH/NAD(+) [347], et de la délétion de la nucléase de surface NucA.

e) **Analyse de perturbations modifiant la composition de la membrane bactérienne**

Ce cinquième projet d'étude de la réponse de *S. agalactiae* en présence d'acide oléique, dans la souche NEM316, ou dans un mutant du gène *fabF* de la voie de biosynthèse des acides gras, a été réalisé en collaboration avec Sophie Brinster et Claire Poyart. Ce projet fait suite à deux études distinctes qui montrent que *S. agalactiae* est capable d'incorporer des acides gras de son environnement et ainsi de palier à un défaut dans la voie de biosynthèse [346, 348]. La première étude indique que la mutation du gène *nox2* diminue le pool d'acétyl-coenzymeA dans la bactérie, et perturbe la biosynthèse des acides gras. Cette perturbation de la biosynthèse des acides gras peut être compensée par l'incorporation d'acide oléique dans les membranes de *S. agalactiae* [346]. La seconde montre que les enzymes intervenant dans la synthèse des acides gras ne sont pas indispensables à *S. agalactiae* à condition que l'environnement fournisse ces acides gras ; ainsi, la mutation du gène *fabF* est létale en TH seul, et peut être compensée par l'ajout d'acide oléique. Cependant le rôle de l'incorporation d'acide oléique dans le cadre de l'interaction de la bactérie avec son hôte reste à être étudié.



## f) **Impact de l'absence d'ancrage des protéines à motif LPXTG**

Ce dernier projet réalisé en collaboration avec Saynoor Dramsi et Mathieu Brochet porte sur l'analyse de l'impact de l'absence d'ancrages des protéines LPXTG, à la surface de la bactérie, sur l'expression génique de NEM316. L'ancrage des protéines à motif LPXTG au peptidoglycane est catalysé par la sortase A [349]. Dans ce projet nous avons étudié l'effet de la délétion ( $\Delta$ srtA) ou de la mutation du site catalytique (srtA\*) du gène codant pour la Sortase A, mais également celui de la mutation de régulateurs impliqués dans l'induction de l'expression de gènes codant pour des protéines à motif LPXTG, comme le régulateur araC (*gbs0469*) qui régule l'expression de la protéine LPXTG Alp2 (*gbs0470*).

### (2) **Résultats**

#### a) **Analyse générale**

Les 34 comparaisons transcriptomiques réalisées dans la souche NEM316 ainsi que le nombre de gènes retrouvés différentiellement exprimés pour chacune des comparaisons sont décrits dans le tableau 8. Ce tableau montre que les différentes conditions testées affectent l'expression d'un nombre variable de gènes. Les mutations affectant le plus grand nombre de gènes sont, par ordre décroissant, la délétion du régulateur transcriptionnel CovR, la mutation du site catalytique de la Ssortase A (SrtA\*), et l'ajout d'acide oléique dans le milieu de culture. Ces trois perturbations conduisent à des modifications importantes du métabolisme de la bactérie et de l'organisation de la paroi bactérienne. Ce grand nombre de gène confirme le rôle central de CovR au sein des réseaux de régulation de *S. agalactiae*. A contrario, la délétion du gène *gbs0469*, codant pour un régulateur de la famille AraC, associé au gène codant pour la protéine Alp2, n'affecte l'expression que d'un nombre limité de gènes. Ce petit nombre de gènes indique que le régulateur Gbs0469 agit probablement en dernière ligne au sein des réseaux de régulation de la souche NEM316.

**Tableau 8: Liste des analyses transcriptomiques réalisées dans la souche NEM316**

Condition1/Condition2	Total <sup>1</sup>	Expression dans		
		Condition1		
		Induite <sup>2</sup>	Réprimée <sup>3</sup>	
<i>ΔdltA</i> /wt	119	112	7	Analyse de la réponse à la <i>ΔdltA</i>
<i>ΔdltR</i> /wt	244	169	75	Analyse de la réponse à la <i>ΔdltA</i>
<i>ΔdltAΔdltR</i> /wt	170	110	60	Analyse de la réponse à la <i>ΔdltA</i>
<i>ΔdltAΔdltR/ΔdltA</i>	73	15	58	Analyse de la réponse à la <i>ΔdltA</i>
<i>ΔciaRH</i> /wt	84	64	20	Analyse de la réponse à la <i>ΔdltA</i>
<i>ΔdltAΔciaRH/ΔdltA</i>	170	110	60	Analyse de la réponse à la <i>ΔdltA</i>
<i>ΔrelRS</i> /wt	123	61	62	Analyse de la réponse à la <i>ΔdltA</i>
<i>ΔdltAΔrelRS/ΔdltA</i>	63	46	17	Analyse de la réponse à la <i>ΔdltA</i>
<i>ΔdltRΔciaRH</i> /wt	118	49	69	Analyse de la réponse à la <i>ΔdltA</i>
Colistine/TH	219	108	111	Réponse suite à la présence de colistine
<i>ΔgraRS</i> /wt	145	48	97	Rôle du système GraRS
<i>Δrga</i> /wt	96	36	60	régulateur de la famille rofA
<i>ΔrogB</i> /wt	178	82	96	régulateur de la famille rofA
<i>Δgbs1426</i> /wt	172	84	88	régulateur de la famille rofA
<i>ΔcovRS</i> /wt	279	195	84	Rôle de AdtA dans le régulon CovRS
<i>ΔcovR</i> /wt	406	242	164	Rôle de AdtA dans le régulon CovRS
<i>ΔcovS</i> /wt	233	174	59	Rôle de AdtA dans le régulon CovRS
<i>ΔadtA</i> /wt	102	41	61	Rôle de AdtA dans le régulon CovRS
Se_v <i>adtA</i> /wt	140	94	46	Rôle de AdtA dans le régulon CovRS
Se_K <i>adtA</i> /wt	232	180	152	Rôle de AdtA dans le régulon CovRS
Se_K <i>gbs1037</i> /wt	316	217	99	Rôle de <i>gbs1037</i> dans les réseaux de régulation
NEM316pTCV/ NEM316	74	59	15	Contrôle
<i>Δrex</i> /wt statique	125	99	26	Réponse lié à l'action de Rex
<i>Δrex</i> /wt aéré	139	104	35	Réponse lié à l'action de Rex
<i>Δgbs0661</i> aéré/wt aéré	185	70	115	Modification d'expression suite a l'absence de NucA
<i>ΔnucA</i> statique /wt statique	139	51	88	Modification d'expression suite a l'absence de NucA
Wt statique /wt aéré	154	98	56	Modification d'expression en culture aéré
<i>ΔnucA</i> statique / <i>ΔnucA</i> aéré	65	30	35	Modification d'expression en culture aéré
TH/ TH OA	312	288	124	Modification de la composition de la membrane
<i>ΔfabF</i> OA/ wt OA	103	35	68	Modification de la composition de la membrane
<i>gbs0469i</i> /wt	16	3	13	Régulateur de la protéine Alpha
<i>ΔsrtA</i> /wt	197	105	92	Réponse suite à l'absence de protéines à motif LPXTG
<i>ΔsrtA*</i> /wt	381	215	166	Réponse suite à l'absence de protéines à motif LPXTG
<i>ΔsrtA*</i> / <i>ΔsrtA</i>	136	50	86	Réponse suite à l'absence de protéines à motif LPXTG

<sup>1</sup>Ce nombre correspond au nombre total de gènes trouvés différentiellement exprimés pour chacun des transcriptomes.<sup>2</sup>Ce nombre indique le nombre de gènes induits dans la condition1 par rapport à la condition 2.

<sup>3</sup>Ce nombre correspond au nombre de gènes réprimés dans la condition1 par rapport à la condition 2.

OA acide oléique Se\_K et Se\_v correspondent à la surexpression de *adtA* ou de *gbs1037* soit au niveau du chromosome soit au niveau du plasmide pTCVerm.

La comparaison des données de 34 analyses transcriptomiques (Tableau 8) montre que sur les 1850 gènes analysés, 629 ont une expression invariable. Parmi ces 629 gènes ne sont pas retrouvés 21 régulateurs simples et 7 TCS (*gbs0121-0122* ; *gbs0309-0310* ; *gbs0429-0430* ; *walKR* ; *gbs1947-1948* ; *gbs2081-2082* ; *gbs2086-2087*), ainsi que deux gènes impliqués dans la transmission du signal: les serine/thréonine phosphatase et kinase *gbs0306* et *gbs0307* (*stp1* et *stk1*). Ces observations suggèrent que ces régulateurs ne sont probablement pas impliqués dans la réponse de *S. agalactiae* aux conditions testées. A contrario l'expression de 19 gènes est affectée dans plus de 50% des analyses (effet supérieur à 2 dans au moins 17 analyses). Ces gènes codent pour: 7 protéines de stress (opéron *gbs1202-1206*, *gbs1778* et *gbs2053*), une protéine associée au ribosome (*gbs0326*), l'alcool déshydrogénase (*gbs0053*), une oxydoréductase (*gbs0196*), le facteur CAMP (*gbs1999-gbs2000*), des systèmes de transport (*gbs1734* et *gbs1779*), une hydrolase et une perméase (PTS) du tréhalose-6-phosphate (*gbs0189-gbs0190*), la protéine à motif LPXTG PilA, une protéine avec un domaine LysM (*gbs2107*), et des protéines de fonctions inconnues (*gbs0600*, *gbs1286*). Cette observation suggère que l'expression de ces protéines est probablement régulée par de multiples réseaux de régulation de *S. agalactiae*, et permet donc de proposer que ces protéines jouent un rôle important dans l'adaptation de *S. agalactiae* à son environnement. Cependant il est également possible que nous ayons biaisé ces résultats en regardant plusieurs mutants appartenant aux mêmes voies de régulation; pour le vérifier il faudrait réaliser d'autres expériences transcriptomiques et regarder si l'on retrouve toujours les mêmes gènes affectés dans plus de 50% des expériences.

## b) **Regroupement des gènes par classification hiérarchique**

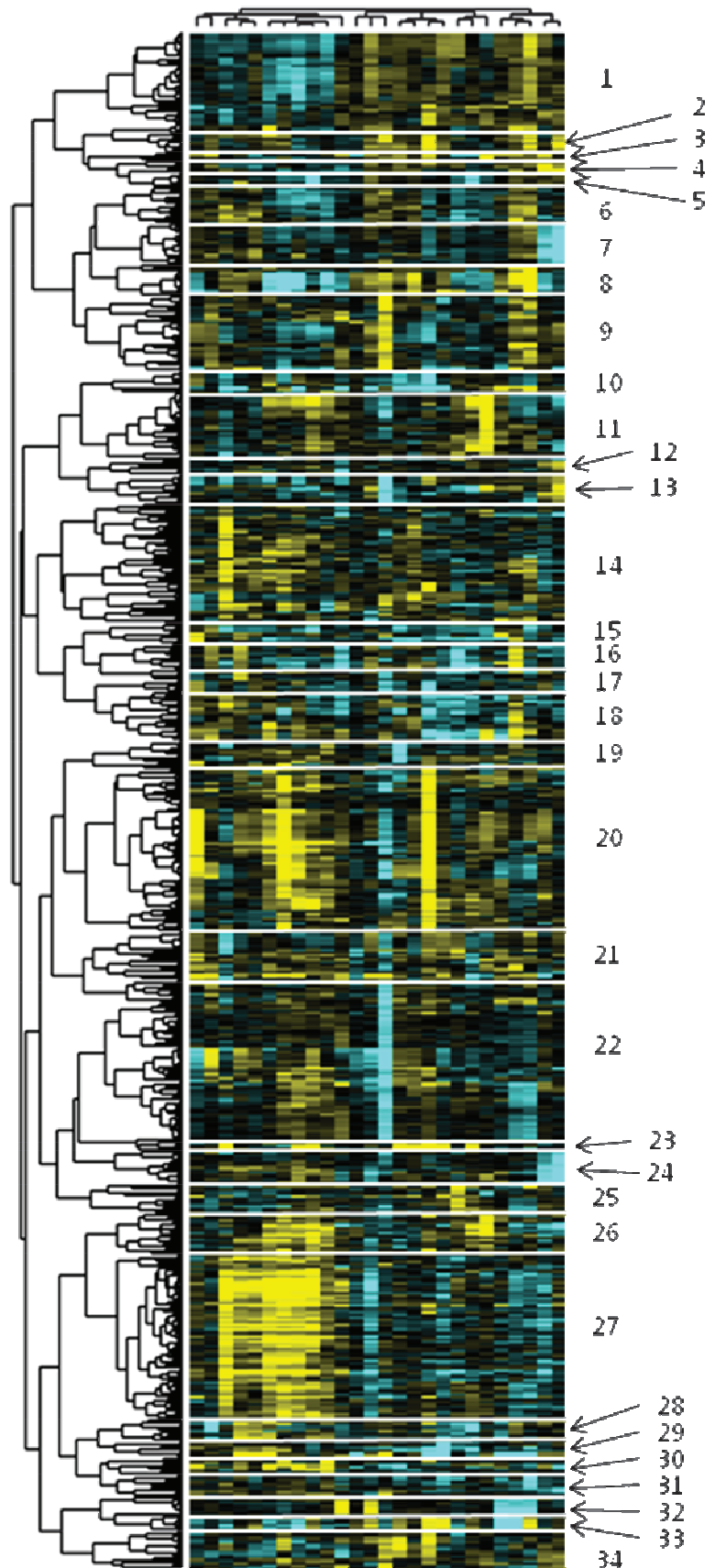
Dans toute approche d'analyse combinée de différentes données transcriptomiques il est recommandé de commencer par une étude exploratoire, en privilégiant les représentations graphiques [350]. Cette étude exploratrice permet de se familiariser avec les données et surtout de dépister les sources de problèmes éventuels, comme les valeurs manquantes, erronées ou atypiques. Suite à cette étude, les données peuvent être prétraitées (retirer les gènes avec des valeurs manquantes, effectuer une transformation logarithmique des données pour limiter l'impact des valeurs extrêmes). Si ces problèmes ne sont pas pris en compte au début de

l'analyse, ils apparaîtront ensuite comme des artefacts susceptibles de fausser les résultats. Pour l'analyse exploratoire deux grandes familles de méthodes sont utilisées: les méthodes factorielles (ACP Analyse en Composantes Principales), et la classification. Suite à cette étude exploratoire des méthodes de modélisation (ANOVA: Analyses de variance et modèle mixte) et des techniques d'apprentissage (K-means) peuvent être utilisées en fonction des questions posées, selon si l'on veut savoir si une condition diffère des autres, ou si l'on souhaite regrouper des gènes au sein de classes mais avec une idée a priori du nombre de classes que l'on souhaite obtenir.

Vu le grand nombre de conditions analysées, j'ai décidé d'utiliser une approche par classification. En tant que première approche, et afin de regrouper les gènes régulés de façon similaire selon les expériences, en différentes classes et sans *a priori* sur le nombre de classes, un clustering hiérarchique a été réalisé. Pour cette classification seules les 26 expériences réalisées versus la souche NEM316, en condition de culture non aérée, ont été utilisées. Cette classification a été réalisée sur les 636 gènes variant d'un facteur 3 dans au moins une des expériences, plutôt que les sur 1233 gènes variant d'un facteur 2 dans au moins une des expériences. La classification obtenue montre que l'on peut regrouper les 636 gènes en 34 groupes différents, pouvant éventuellement être expliqués par l'action d'un régulateur transcriptionnel. Je vais présenter une partie de ces groupes dans le paragraphe suivant.

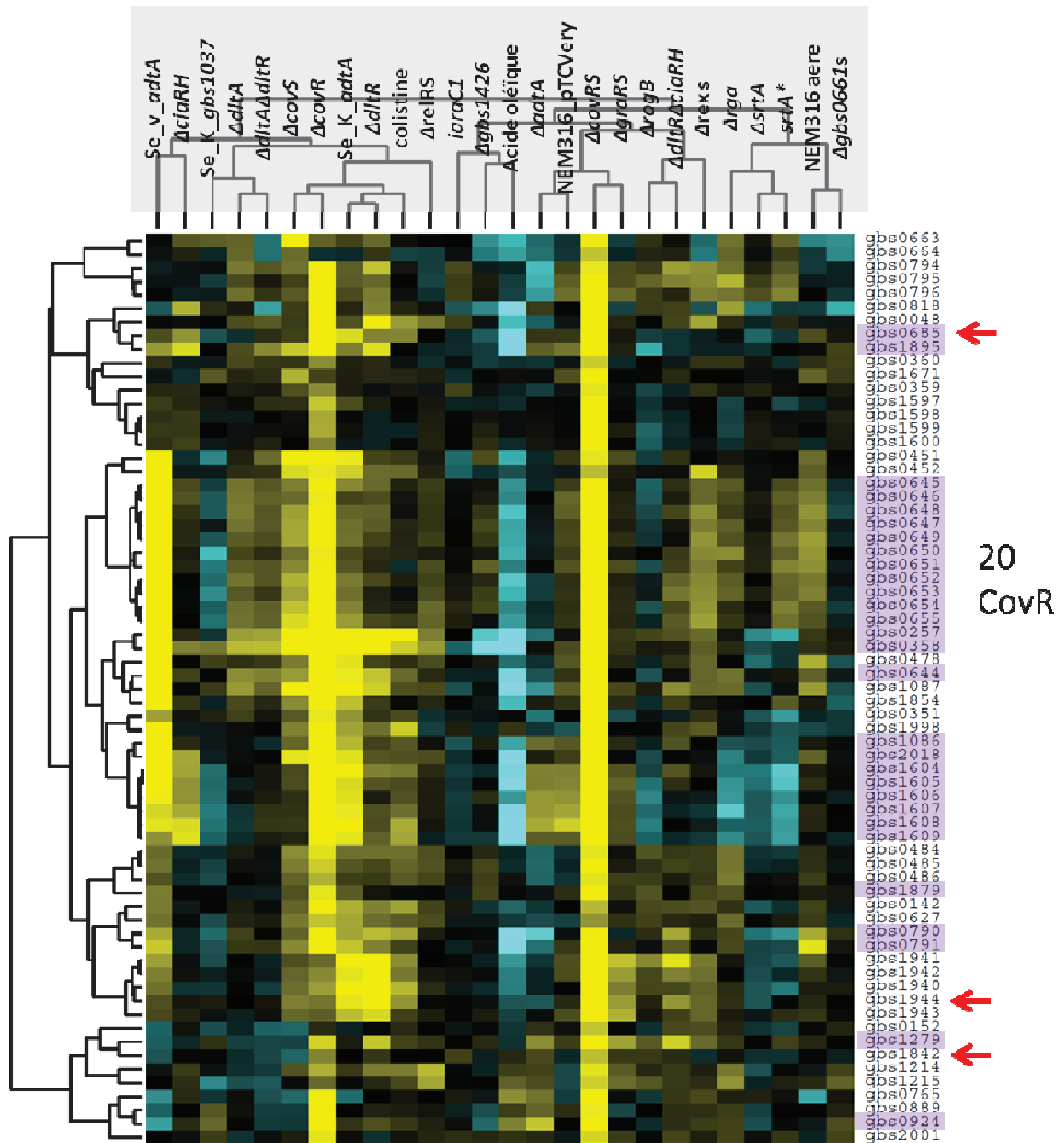
**Figure 21: Classification hiérarchique de 636 gènes.**

Liste des comparaisons: (gauche à droite): Se\_v\_ *adtA* /wt;  $\Delta$ *ciaRH*/wt; Se\_K\_ *gbs1037*/wt;  $\Delta$ *dltA*/wt;  $\Delta$ *dltA $\Delta$ *dltR*/wt;  $\Delta$ *covS*/wt;  $\Delta$ *covR*/wt; Se\_K\_ *adtA* /wt;  $\Delta$ *dltR*/wt; colistine/wt;  $\Delta$ *relRS*/wt; *iaraC1*/wt;  $\Delta$ *gbs1426*/wt; NEM316 OA/NEM316;  $\Delta$  *adtA* /wt; NEM316pTCVerm/NEM316;  $\Delta$ *covRS*/wt;  $\Delta$ *graRS*/wt;  $\Delta$ *rogB*/wt;  $\Delta$ *dltR $\Delta$ *ciaRH*/wt;  $\Delta$ *rexs*/wts  $\Delta$ *rga*/wt;  $\Delta$ *srtA*/wt; *srtA\**/wt; NEM316a/NEM316;  $\Delta$ *gbs0661s*/wts  
 Les abréviations OA, s, a correspondent respectivement à l'acide oléique, à statique, et à aéré. Se\_K et Se\_v correspondent à la surexpression de *adtA* ou de *gbs1037* soit au niveau du chromosome soit au niveau du plasmide pTCVerm. Les gènes induits sont indiqués en jaunes et les gènes réprimés en bleu.**



### c) **Analyse des différents groupes de gènes**

Dans un premier temps afin de valider la classification obtenue, je me suis intéressée aux cibles connues des trois régulateurs CovRS, CiaRH et Rga. L'analyse du clustering montre que deux des groupes, 28 et 33, correspondent à des gènes régulés par CiaRH et Rga. Ces groupes sont respectivement composés de 9 et 7 gènes et ont été précédemment décrits dans les parties UU.1 et UU.2. Les 29 gènes correspondant à des cibles directes identifiées de CovRS sont regroupés dans le cluster 20 [230] (indiqués en violet sur la figure 22). Notamment on retrouve parmi ces gènes l'opéron *cyl* (*gbs0644* à *gbs0655*), les gènes codant pour les protéines LPXTG Gbs2018 et Gbs0791, et les gènes de fonctions inconnus (*gbs0257*, *gbs0358*). Outre ces 29 gènes le cluster 20 contient 38 autres gènes dont trois régulateurs transcriptionnels qui pourraient constituer de nouvelles cibles de CovR. L'ensemble de ces observations valide nos résultats de classification hiérarchique.

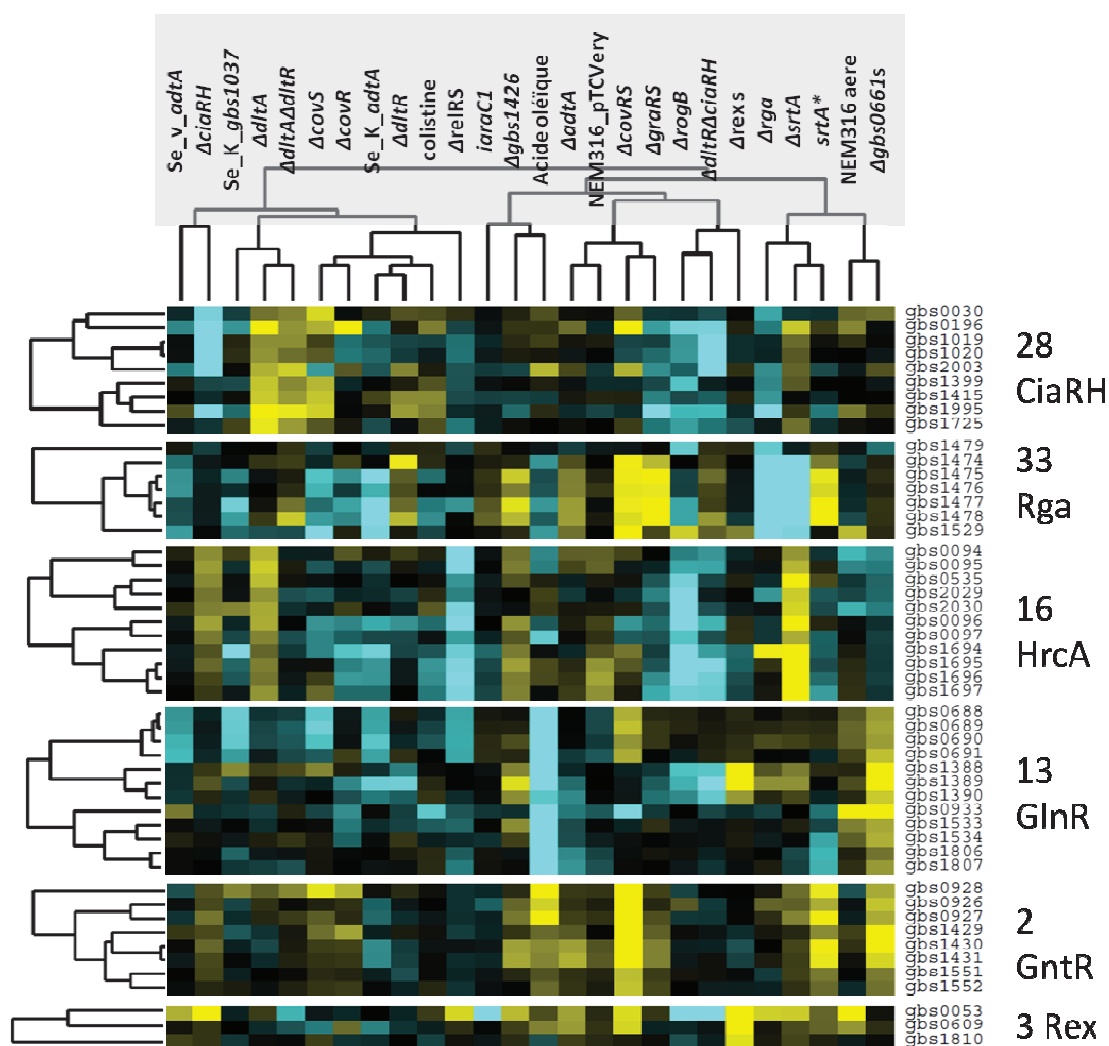


**Figure 22: Le cluster CovR.**

Les abréviations OA, s, a correspondent respectivement à l'acide oléique, à statique, et à aéré. Se\_K et Se\_v correspondent à la surexpression de *adtA* ou de *gbs1037* soit au niveau du chromosome soit au niveau du plasmide pTCVerm. Les gènes colorés en violet correspondent aux gènes appartenant au régulon conservé de CovR. Et les gènes indiqués par des flèches rouges correspondent à des régulateurs transcriptionnels. Les gènes induits sont indiqués en jaunes et les gènes réprimés en bleu.

Dans un second temps, j'ai tenté d'identifier les régulateurs potentiellement impliqués dans les différents groupes de régulation. Pour cela, je me suis basée sur l'observation du fait que la majorité des régulateurs transcriptomiques s'autorégulent [351, 352]. L'existence de telles boucles de régulation a pour conséquence que les gènes codant pour des régulateurs transcriptionnels sont souvent co-régulés avec les gènes qu'ils ciblent, et donc regroupés avec leurs cibles lors de la classification hiérarchique [323]. J'ai donc recherché la présence de régulateurs au sein des clusters. Ces observations ont alors été confrontées aux données bibliographiques sur ces régulateurs. Parmi les 636 gènes étudiés, 35 codent pour des régulateurs transcriptionnels et 6 pour des histidines kinases de systèmes à deux composants (Annexes TableS 2).





**Figure 23: Les clusters CiaRH, Rga, HrcA, GlnR, GntR et Rex.**

Les abréviations OA, s, a correspondent respectivement à l'acide oléique, à statique, et à aéré. Se\_K et Se\_v correspondent à la surexpression de *adtA* ou de *gbs1037* soit au niveau du chromosome soit au niveau du plasmide pTCVerm. Les gènes coloré en violet correspondent aux gènes appartenant au régulon conservé de CovR. Et les gènes indiqués par des flèches rouges correspondent à de régulateurs transcriptionnels. Les gènes induits sont indiqués en jaunes et les gènes réprimés en bleu.

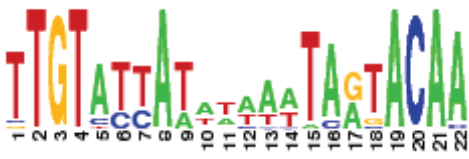
La localisation systématique des 35 régulateurs (indiquées par une flèche rouge) de la Table S 2 dans les différentes classes a permis d'identifier quatre autres régulons potentiels.

Le groupe n° 2 contient huit gènes répartis en trois unités transcriptionnelles dont un gène code pour un régulateur transcriptionnel de la famille GntR (Gbs0928) (Figure 21 et 23). Cette

famille a été nommée GntR après l'identification du répresseur de l'opéron gluconate chez *B. subtilis* [353]. Les sept autres gènes codent pour trois ABC transporteurs (gbs0927-26, gbs1431-29, gbs1551-52). Chez *S. pneumoniae* il a été démontré que l'homologue de Gbs0928 (SP1714) agissait comme répresseur sur sa propre expression et sur l'expression de deux ABC transporteurs (SP1715, et SP0785-87), qui interviennent dans la résistance de *S. pneumoniae* aux peptides antimicrobiens [336]. Parmi les huit gènes identifiés dans le groupe n° 2, cinq sont des orthologues des gènes identifiés chez *S. pneumoniae*, ce qui suggère que GntR pourrait aussi réguler l'expression de ces gènes chez *S. agalactiae*. Les régulateurs appartenant à la famille GntR au sein du genre streptocoque se fixent au niveau de séquences palindromiques TTGTATTA-N6-TAGTACAA [354, 355]. Ce site est retrouvé en amont de deux des opérons codant pour ces ABC transporteurs. Un site moins caractéristique en 5' est également retrouvé au niveau de la séquence -10 du promoteur du gène *gbs1552* (Table 8). On constate que les sites de fixation de GntR chevauchent la boîte -10 prédite lors des expériences d'identifications des sites d'initiation de la transcription (séquence soulignée dans le Tableau 8).

Ce groupe de gènes a été identifié alors qu'aucune expérience ne portait sur un de ses constituants. L'expression des gènes du groupe n° 2 est induite lors de la délétion du système CovRS, ou de la nucléase de surface NucA, et par la mutation du site catalytique de la Sortase A. Cependant, il semblerait que cette modification de l'expression soit une conséquence indirecte, et probablement indépendante de CovRS ou de SrtA car l'expression de ces gènes n'est pas retrouvée altérée dans d'autres mutants comme  $\Delta covS$ ,  $\Delta covR$ , et  $\Delta srtA$ .

**Tableau 9: Identification de sites de fixation potentiels de GntR**

	Distance par rapport au +1 (ATG)		Conservation du site de fixation potentiel au sein des streptocoques
<i>gbs0928-26</i>	+7(-30)	<b>TTGTA</b> CTATttttaa <b>TAGTACA</b> A	Pn, Py, Mu
<i>gbs1431-29</i>	+15(-14)	taata <b>cac</b> TTGTt <b>CTAT</b> attaa <b>TAATA</b> CAA	Pn, Py, Mu
<i>gbs1552</i>	-5 (-33)	ttaaaccacagtaa <b>TAGTACA</b> ATgatataaaggt	


La séquence consensus du site de fixation de GntR est indiquée en LOGO [354, 355]. Le tableau indique pour chaque unité de transcription, les séquences nucléiques correspondant au site potentiel de GntR chez la souche NEM316, ainsi que la distance en nucléotides de ce site par rapport au +1 de transcription. Les séquences correspondant à la boîte -10 sont soulignées. Les abréviations Pn, Py et Mu permettent d'indiquer si la présence des sites de fixation de GntR est également observée en amont des orthologues respectivement de ces gènes chez *S. pneumoniae*, *S. pyogenes* et *S. mutans*.

Le groupe n°13 constitué de 12 gènes organisés en 5 unités transcriptionnelles (figure 21 et 23), regroupe des gènes dont l'expression est réprimée par la croissance en acide oléique. Au sein de ces gènes est retrouvé un gène codant pour un régulateur transcriptionnel appartenant à la famille MerR et de type GlnR (*gbs1807*). Chez *S. mutans*, il a été démontré, que ce régulateur intervient dans la répression de gènes impliqués dans la synthèse et le transport de la glutamine et du glutamate en condition de stress acide [356]. Le site de fixation de GlnR, d'abord identifié chez *B. subtilis* [357], ATGTNA-N7-TNACAT semble conservé au sein des espèces bactériennes à Gram positif [358]. Chez *S. agalactiae* ce site est présent en amont de 3 opérons (tableau 10) codant pour 6 protéines impliquées dans la régulation (*glnR*), la synthèse (*glnA*) et le transport (*glnPQ* et *glnQHMP*) de la glutamine et du glutamate. Cependant, aucun site potentiel de fixation de GlnR n'est observé au niveau des promoteurs de l'opéron *gbs1388-89* et du gène *gbs0933* également localisés dans ce cluster, qui codent respectivement pour deux protéines de fonctions inconnues, et pour la glucosamine-fructose-6-phosphate aminotransférase, également impliquée dans le métabolisme de la glutamine.

L'expression des gènes du groupe GlnR (n°13 figure 23) est réprimée par l'ajout d'acide oléique dans le milieu de culture, et est induite par la délétion de la nucléase de surface NucA, suggérant que les besoins de *S. agalactiae* en glutamine et glutamate sont diminués en présence d'acide gras dans le milieu de culture. Cette observation peut être potentiellement liée aux besoins énergétiques moins importants de *S. agalactiae* en présence d'acide oléique, la voie de

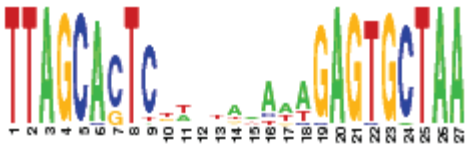
synthèse des acides gras étant réprimée par l'ajout d'acide oléique ou la présence de sérum dans le milieu [348].

**Tableau 10: Identification de sites de fixation potentiels de GlnR**

	Distance par rapport au +1 (ATG)	 ATGTNA-N7-TNACAT	Conservation du site de fixation potentiel au sein des streptocoques
glnPQ ( <i>gbs1533-34</i> )	-52 (-74)	ATGTTAcgttttcTAACAT	Pn,Py,Mu
glnRA ( <i>gbs1807-06</i> )	-43 (-56)	ATGTCAtaaaatcTCACAT	Pn,Py,Mu
glnQHMP ( <i>gbs0691-88</i> )	+20 (-25)	ATGTTAggtaaccTAACAa	Pn,Mu

La séquence consensus du site de fixation de GlnR est indiquée en LOGO [358]. Le tableau indique pour chaque unité de transcription, les séquences nucléiques correspondant au site potentiel de fixation de GlnR chez la souche NEM316, ainsi que la distance en nucléotides de ce site par rapport au +1 de transcription, et par rapport au codon d'initiation. Les abréviations Pn, Py et Mu permettent d'indiquer si la présence des sites de fixation de GlnR est également observée en amont des orthologues de ces gènes respectivement chez *S. pneumoniae*, *S. pyogenes* et *S. mutans*.

Le gène codant le régulateur HrcA est localisé dans le cluster n° 16 constitué de onze gènes organisés en 4 opérons. Ce régulateur, d'abord identifié chez *B. subtilis*, est impliqué dans la régulation de l'expression de chaperonnes impliquées dans la résistance au choc thermique [359, 360]. Ce régulateur est retrouvé chez des bactéries Gram positif, mais également chez des protéobactéries et des cyanobactéries [361, 362]. Au sein du genre streptocoque ce régulateur a été étudié chez *S. pneumoniae* [363, 364], *S. thermophilus* [365], *S. mutans* [366, 367] et *S. pyogenes* [368]. Le régulateur HrcA se fixe au niveau de séquences palindromiques (TTAGCACTC-N9-GAGTGCTAA) appelées CIRCE (pour "controlling inverted repeat of chaperone expression") [359, 360]. Trois des opérons mis en évidence semblent directement régulés par HrcA, un site CIRCE étant identifiable au niveau de leur séquence promotrice (Table 11). Nos résultats combinés avec l'ensemble des études réalisées au sein du genre streptocoque indiquent que deux des opérons cibles *hrcA-grpE-dnaKJ* et *groSL* sont conservés au sein des streptocoques. Cependant le gène *clpE* semble régulé par HrcA uniquement chez *S. agalactiae*; en effet, aucun site CIRCE n'est identifiable en amont de ce gène chez *S. pneumoniae*, *S. pyogenes*, et *S. mutans*.


	Distance par rapport au +1 (ATG)		Conservation du site de fixation potentiel au sein des streptocoques
<i>hrcA-grpE-dnaKJ</i>	+31 (-32)	TTAGCAGTCGATGTGTAAGAGTGCTAA	Pn,Py,Mu
	-52 (-114)	TTAGCACTCTTTGAATAAGAGTGCTAA	Pn,Py,Mu
<i>clpE</i>	-51 (-117)	TTAGCACTCTAATATATTCAGTGCTAA	
<i>groSL</i>	+31 (-38)	TTAGCACTCTCTGTATAGGAGTGCTAA	Pn,Py,Mu

**Tableau 11: Identification de sites de fixation potentiels de HrcA**

La séquence consensus du site CIRCE est indiquée en LOGO [359, 360]. Le tableau indique pour chaque unité de transcription, les séquences nucléiques correspondant au site potentiel de HrcA chez la souche NEM316, ainsi que la distance en nucléotides de ce site par rapport au +1 de transcription, et par rapport au codon d'initiation. Les abréviations Pn, Py et Mu permettent d'indiquer si la présence des sites de fixation de HrcA est également observée en amont des orthologues de ces gènes chez *S. pneumoniae*, *S. pyogenes* et *S. mutans* respectivement.

Le groupe n°3 correspond à trois gènes dont l'expression est induite chez le mutant du régulateur *rex* ( $\Delta gbs1167$ ). Le gène *rex* (pour redox sensing) code pour un régulateur transcriptionnel dont l'activité chez *B. subtilis* dépend du ratio NADH/NAD(+) [347]. Au sein de ce cluster est retrouvé le gène *gbs0053* codant pour l'alcool-acétaldéhyde déshydrogénase. Or, chez *S. mutans* et *S. aureus*, le régulateur Rex réprime l'expression de l'alcool-acétaldéhyde déshydrogénase et son action directe a été confirmée chez *S. aureus* [369]. La séquence du site de liaison de Rex chez *S. aureus* a été identifiée et correspond aux séquences palyndromiques TTGTGAA-4nt-TTCACAA [369]. Chez *S. agalactiae* ce site est retrouvé en amont des gènes *gbs0053*, *gbs0054*, et du gène *gbs0609* appartenant au cluster n°3. Un second cluster (n°11) correspond également à des gènes dont l'expression est induite dans le mutant *rex*. Cependant aucun site n'est identifiable en amont de ces gènes, suggérant que leur régulation est indirecte et fait probablement intervenir un autre régulateur.

**Tableau 12: Identification de sites de fixation potentiels de Rex**

		 TTGTGAA-N4-TTCACAA	Conservation du site de fixation potentiel au sein des streptocoques
<i>gbs0053</i>	+28 (-18) -28 (-73) -51 (-96)	TTGTGAAACATTTAACAA TTGTGAAACTATTACAA TTGGAAAAAAGTTCACAA	Mu, Pn Mu, Pn Mu, Pn
<i>gbs0054</i>	+16 (-32) -9 (-56) -54 (-104)	ATGTAAACGATTACAAA TTGTGAAGAATAATACAA TTAAGAAATATTTCAAAA	Py Pn
<i>gbs0609*</i>	- 58 -83	TTCAGAAATGCATTTAAAA TTGTTACAAAAATCACAT	

La séquence consensus du site de liaison de Rex est indiquée en LOGO [369]. Le tableau indique pour chaque unité de transcription, les séquences nucléiques correspondant au site potentiel de Rex chez la souche NEM316, ainsi que la distance en nucléotides de ce site par rapport au +1 de transcription, et par rapport au codon d'initiation. Les abréviations Pn, Py et Mu permettent d'indiquer si la présence des sites de fixation de Rex est également observée en amont des orthologues de ces gènes respectivement chez *S. pneumoniae*, *S. pyogenes* et *S. mutans*. \*Pas de site d'initiation décrit pour ce gène.

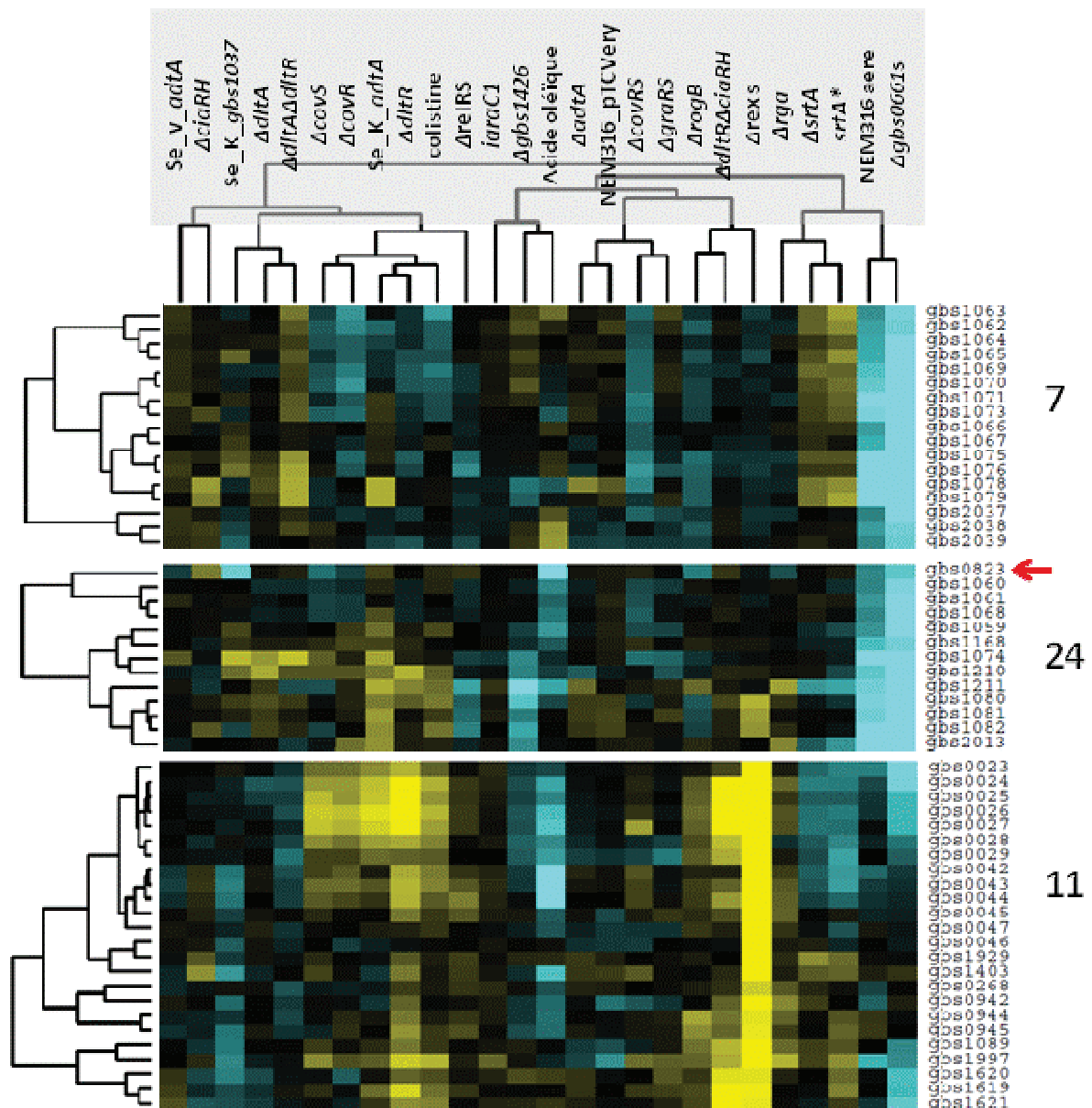
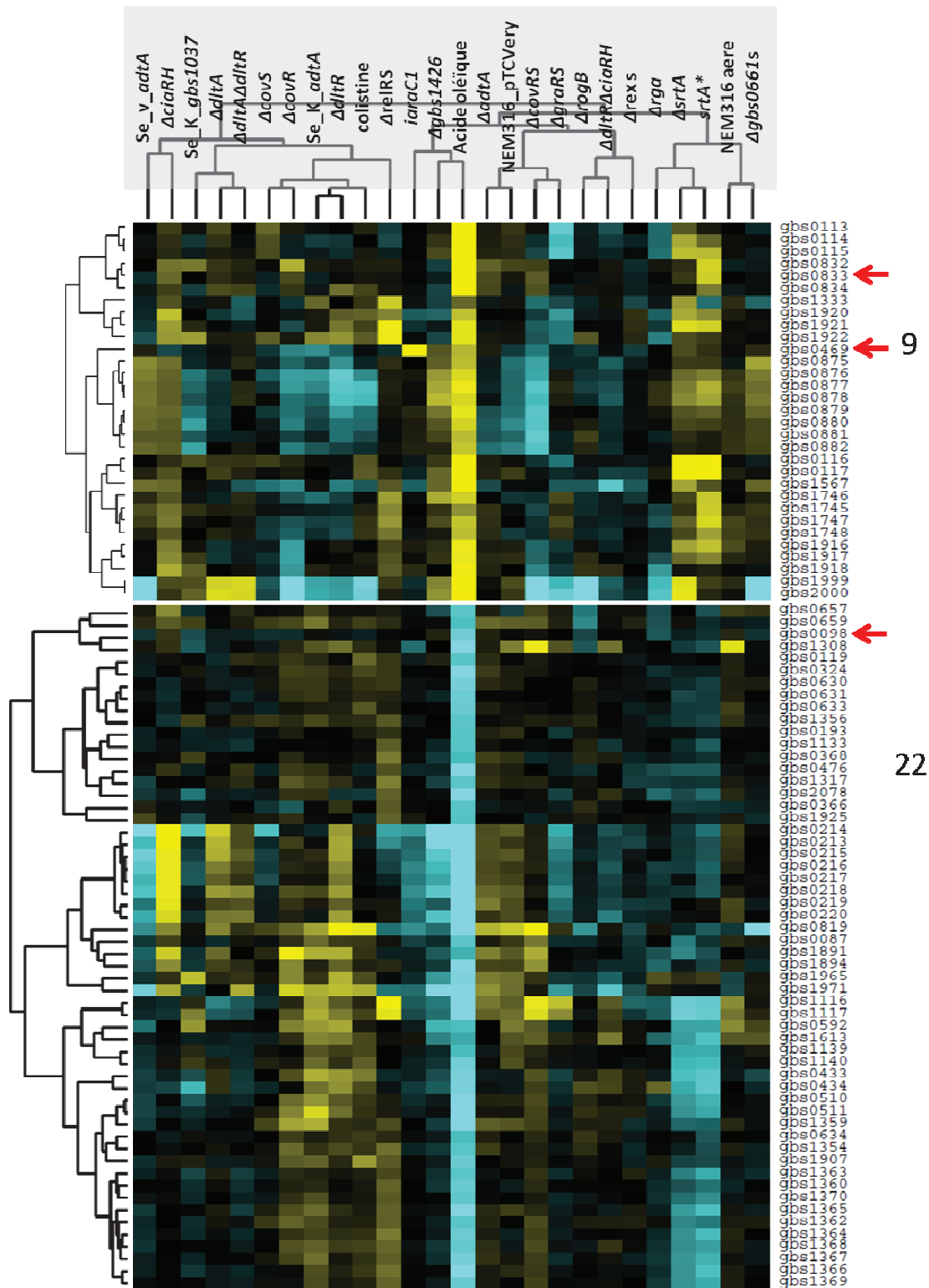


Figure 24: Les clusters 7, 24 et 11.

Les abréviations OA, s, a correspondent respectivement à l'acide oléique, à statique, et à aéré. *Se\_K* et *Se\_v* correspondent à la surexpression de *adtA* ou de *gbs1037* soit au niveau du chromosome soit au niveau du plasmide pTCVerm. Les gènes colorés en violet correspondent aux gènes appartenant au régulon conservé de CovR. Et les gènes indiqués par des flèches rouges correspondent à des régulateurs transcriptionnels. Les gènes induits sont indiqués en jaunes et les gènes réprimés en bleu.



**Figure 25: Clusters 9 et 22.**

Les abréviations OA, s, a correspondent respectivement à l'acide oléique, à statique, et à aéré. Se\_K et Se\_v correspondent à la surexpression de *adtA* ou de *gbs1037* soit au niveau du chromosome soit au niveau du plasmide pTCVerm. Les gènes colorés en violet correspondent aux gènes appartenant au régulon conservé de CovR. Les gènes indiqués par des flèches rouges correspondent à des régulateurs transcriptionnels. Les gènes induits sont indiqués en jaunes et les gènes réprimés en bleu.



Les gènes répondant à la présence d'acide oléique dans le milieu sont regroupés majoritairement au sein de deux groupes: le groupe n°9 réunit des gènes induits par l'acide oléique et le groupe n°22 contient à l'inverse des gènes réprimés. Il y a parmi les gènes du groupe 9 deux régulateurs transcriptionnels correspondant respectivement au régulateur de la famille *AraC* (*gbs0469*), et un régulateur de type Cro/CI (*gbs0833*). Le cluster 22 comprend un gène codant pour un régulateur transcriptionnel de la famille GntR (*gbs0098*). Parmi les gènes regroupés au sein du groupe n° 22, 24 correspondent à des gènes localisés au sein d'EGMs, tels que les gènes *gbs0214* à *gbs0220*, et les trois gènes appartenant à TnGBS2 (*gbs1133*, *gbs1139* et *gbs1140*), et des gènes localisés entre *gbs1309* et *gbs1373* qui correspondent à l'îlot génomique XII, suggérant que l'expression des EGMs de *S. agalactiae* est réprimée par la présence d'acide oléique dans le milieu de culture.

Les clusters 7 et 24 regroupent des gènes dont l'expression est majoritairement réprimée par la croissance en condition d'aération ou par la délétion de la nucléase de surface. Le cluster 7 est constitué de 17 gènes et ne contient aucun régulateur transcriptionnel. L'ensemble des gènes identifiés dans ce cluster sont répartis dans deux régions chromosomiques, dont une correspond à l'îlot génomique IX (de *gbs1053* à *gbs1082*) [39]. L'autre regroupe trois gènes *gbs2037*, *gbs2038* et *gbs2039*, codant respectivement pour une réductase de ribonucléotides en condition anaérobie, une acétyltransférase, et une oxidoréductase. Le cluster 24 contient 13 gènes dont 9 appartiennent également à l'îlot génomique IX. Au sein de ce cluster est également trouvé un gène (*gbs0823*) codant pour une protéine avec un domaine CoiA (pfam06054). Les protéines de cette famille sont souvent annotées comme des régulateurs transcriptionnels en raison de la présence d'un domaine en doigt de zinc dans la partie N-terminale. Cependant, ces protéines agiraient comme endonucléase, dans le processus de compétence après l'entrée de l'ADN [370, 371]. Les autres gènes *gbs1168*, *gbs1210*, *gbs1211* et *gbs2013* codent respectivement pour la protéine impliquée dans la réparation de l'ADN RadC, une protéine impliquée dans la dégradation de l'acide phénylacétique, une uracil perméase, et un transporteur de la famille MFS (pour "Major facilitator Superfamily").

Le groupe 34 regroupe 62 gènes dont l'expression est induite dans les mutants  $\Delta dltA$ ,  $\Delta dltRA$ ,  $\Delta covS$ ,  $\Delta covR$ , et  $\Delta dltR$ , ainsi que dans les mutants de surexpression de *adtA* et de *gbs1037* (un autre gène codant pour une protéine avec un domaine Abi). Au sein de ces gènes

quatre régulateurs transcriptionnels sont identifiables. Cette observation suggère que les gènes de ce cluster semblent régulés par ces différents systèmes au sein des mêmes réseaux de régulation.

D'autres groupes de gènes tels que les cluster n°1, 8 et 12 regroupent des gènes dont l'expression est liée à la croissance ou au stress de la bactérie. Par exemple au sein des 42 gènes du cluster n°1, 14 gènes codent pour des protéines ribosomales, pour la sous unité alpha de l'ARN polymérase, pour la protéine de réparation RecN, et pour les protéines du locus CRISPR, et 5 pour des protéines de la voie de respiration aérobie dont les gènes *cydABCD*. Au sein du cluster 8 sont retrouvés 7 gènes codant pour des protéines du système de transport et de biosynthèse de l'arginine, qui est connu pour être impliqué dans la résistance aux stress acide chez de nombreuses espèces bactériennes [372, 373]. Le groupe n°12 contient le gène *gbs1426* codant pour le troisième régulateur de la famille RofA like et également 5 autres gènes organisés en 4 unités transcriptionnelles et codant respectivement pour deux protéines ribosomiques RpmA et RplU (*gbs1438*, et *gbs1440*), un facteur d'élongation de la transcription GreA (*gbs1661*), et deux protéines de fonctions inconnues (*gbs1439* et *gbs1218*).

L'opéron *dltRSABCD* forme un cluster distinct le cluster n°5, l'expression des gènes de l'opéron *dlt* n'est pas modifiée dans les conditions réalisées autres que celles abordées dans la partie UU.1.

L'analyse comparée des différents transcriptomes permet de prédire les cibles de sept régulateurs transcriptionnels de *S. agalactiae*. Cette analyse comparée m'a permis de caractériser des régulons qui n'ont pas été étudiés en tant que tels, comme les régulons GntR, et GlnR, alors que la majorité de ces expériences ont été réalisées en utilisant le milieu Todd Hewitt. Certaines cibles de ces régulateurs sont conservées au sein d'autres espèces du genre streptocoque. Ces analyses constituent une sorte de preuve de l'approche qui ne m'a néanmoins pas permis d'expliquer l'ensemble des groupes de gènes de la classification. En effet, il existe différentes limitations possibles à l'approche utilisée:

- J'ai fait l'hypothèse que le ou les régulateurs impliqués dans la régulation des gènes d'un groupe étaient autorégulés et appartenaient donc à ce cluster. Ce n'est clairement pas systématique.

- Pour des raisons pratiques je me suis basée pour cette classification sur les gènes différentiellement exprimés d'un ratio de 3 dans une des expériences ; or, beaucoup de gènes codant pour des régulateurs transcriptionnels varient avec des ratios inférieurs à 3.
- Pour étudier ces régulons je me suis basée sur les données bibliographiques des régulateurs orthologues au sein du genre streptocoque.

Cette analyse m'a également permis de prédire d'autres régulateurs en plus de Rga potentiellement régulés par CovR. Leur analyse pourrait permettre d'expliquer une autre partie du régulon CovR de *S. agalactiae*. Cette analyse pilote suggère également des interconnexions entre les différents régulateurs et permet une première analyse de la structure des réseaux de régulation chez *S. agalactiae*.

## II. Etude d'une nouvelle famille d'ICE, les TnGBS

L'étude du génome de la souche NEM316 a mise en évidence la présence de deux EGMs particuliers: TnGBS1 et TnGBS2, présents respectivement en trois et une copies. Chaque élément code pour des protéines potentiellement impliquées dans le transfert par conjugaison et dans la réplication, ce qui leurs avaient valu, à l'époque de l'annotation du génome de la souche NEM316, le nom de "plasmides intégratifs" [125]. Par la suite, ces éléments ont été assimilés à des éléments conjugatifs et intégratifs (ICE), leurs extrémités étant constituées de séquences répétées inversées (IRs), flanquées de répétitions directes, caractéristiques de l'insertion par un mécanisme de transposition. L'analyse du génome de nombreuses souches par une approche d'hybridation sur puce à ADN, et par séquençage, montre la structure composite du génome de *S. agalactiae*, avec un squelette conservé et 11 à 14 grandes régions variables entre les souches [39, 234]. TnGBS1 et TnGBS2 correspondent respectivement aux îlots III, VII, VIII pour TnGBS1 et à l'îlot X pour TnGBS2 de la souche NEM316. Cette approche par hybridation sur puce à ADN, réalisée à partir de la séquence génomique de la souche NEM316, a permis de constater que ces éléments, ou des éléments apparentés, étaient retrouvés au sein du génome d'autres souches de *S. agalactiae*. Les présences de TnGBS1, de TnGBS2, et d'éléments apparentés au sein des génomes d'autres souches, suggèrent que ces transposons peuvent être transmis horizontalement entre souches de *S. agalactiae*. La seconde partie de ma thèse porte sur l'étude de la diversité des éléments de cette nouvelle famille d'ICEs, étude qui a conduit à l'identification d'éléments pouvant intervenir dans la régulation du processus de conjugaison. Dans un premier temps, je rapporterai les résultats obtenus sur la validation fonctionnelle de la mobilité de ces éléments, et tout particulièrement de leurs excision et intégration, puis l'analyse de la diversité des TnGBS et la comparaison des modules fonctionnels portés par ces éléments. Enfin je présenterai des données préliminaires sur la régulation du processus de conjugaison et les éléments potentiellement impliqués dans cette régulation.

### A. Matériels et Méthodes

#### 1. Culture des souches

Les souches de *S. agalactiae*, *S. gallolyticus*, *S. macedonicus*, et *E. faecalis* ont été mises en culture en milieu TH, alors que les souches de *E. coli* ont été mises en culture en milieu LB.

Les antibiotiques ont été utilisés aux concentrations suivantes. Pour l'érythromycine, 150µg/mL pour *E. coli* et 10µg/mL pour *S. agalactiae*, *S. gallolyticus*, *S. macedonicus*, et *E. faecalis*. Pour la tétracycline, 10µg/mL pour *S. agalactiae* et *S. gallolyticus*.

## 2. Construction des mutants

Les mutants de délétion des gènes *gbs1147* et *gbs1141* ont été construits en utilisant la méthode de PCR chevauchante décrite précédemment par Shaynoor Dramsi [85]. Une première série de PCR a été réalisée avec les oligonucléotides 1141\_*EcoRI*-1141\_Int1 et 1141\_Int2-1141\_*BamHI* pour le gène *gbs1141*, et de la même manière avec les oligonucléotides 1147\_*EcoRI*-1147\_Int1 et 1147\_Int2-1147\_*BamHI* pour le gène *gbs1147*. Les produits de ces PCR ont ensuite servis de matrice pour une seconde série de PCR, en utilisant les oligonucléotides 1141\_*EcoRI*-1141\_*BamHI* et 1147\_*EcoRI*-1147\_*BamHI*. Les produits PCR ainsi obtenus ont été clonés dans le plasmide thermosensible pG1, digéré par les enzymes *EcoRI* et *BamHI*. Les plasmides recombinants résultant ont été électroporés dans la souche NEM316 de *S. agalactiae*. Les transformants et intégrants ont été sélectionnés comme décrit dans la première partie. Pour isoler des mutants de délétion des gènes *gbs1141* et *gbs1147*, les intégrants sont cultivés par passages successifs en milieu TH, sans érythromycine et à 30°C, ce qui permet l'excision du plasmide. Des clones, isolés à partir de ces cultures successives, ont été testés, d'une part par PCR, d'autre part pour leur sensibilité à l'érythromycine. Les délétions des gènes *gbs1141* et *gbs1147* ont été vérifiées par PCR puis séquençage, en utilisant respectivement les oligonucléotides 1141verifF-1141verifR et 1147verifF-1147verif R.

**Tableau 13: Liste des primers utilisés pour la construction des mutants et des différents plasmides utilisés dans la 2ème partie de la thèse**

Nom oligos	Séquence	
<b>1141_EcoRI</b>	CTAGAATTCTCCTCTCATAATCGGGTTGC	mutant $\Delta gbs1141$
<b>1141_Int1</b>	GGCACGCCCCGGGTGCTGCCGCGGCAAATGCCACATAATCAA	mutant $\Delta gbs1141$
<b>1141_Int2</b>	GCGGCAGCACCCGGGCGTGCCATTTCTCAGTATTTTAGCCTTAGTCA	mutant $\Delta gbs1141$
<b>1141_BamHI</b>	CAGGATCCGTTGCCCAATTAGATACGC	mutant $\Delta gbs1141$
<b>1147_EcoRI</b>	CTAGAATTCCAGGAATTGGAACGTGAGGT	mutant $\Delta gbs1147$
<b>1147_Int1</b>	GGCACGCCCCGGGTGCTGCCGCTGTCAAGTGGAGTACCCCCCTTC	mutant $\Delta gbs1147$
<b>1147_Int2</b>	GCGGCAGCACCCGGGCGTGCCACGGATAATGCTAAAAAGCTCTCA	mutant $\Delta gbs1147$
<b>1147_BamHI</b>	CAGGATCCTAGCACACCTTGCTTTCCT	mutant $\Delta gbs1147$
<b>1141verifF</b>	CACCAGCTTTACCATCCACA	vérification du mutant $\Delta gbs1141$
<b>1141verifR</b>	ATCCAGAAGGCCACAATGAC	vérification du mutant $\Delta gbs1141$
<b>1147verifF</b>	TCTACCAAGCTAAGCGTCAGG	vérification du mutant $\Delta gbs1147$
<b>1147verifR</b>	CCATCCACATAGCCTTAGCA	vérification du mutant $\Delta gbs1147$
<b>P</b>	CGCCAGGGTTTTCCAGTCACGAC	
<b>RP</b>	AGCGGATAACAATTTACACAGGA	
<b>P<sub>IRT2</sub>_EcoRI</b>	CAGGAATTCATGTCGTGAACCACTATCCT	Promoteur P <sub>IR</sub>
<b>P<sub>IRT2</sub>_BamHI</b>	CAGGATCCCCTTCACTAGTTTTGACCAAC	Promoteur P <sub>IR</sub>
<b>P<sub>IRT2</sub>inv_EcoRI</b>	AACTAGAATTCCTTCACTAGTTTTGACCAAC	Promoteur P <sub>IR</sub> inv
<b>P<sub>IRT2</sub>inv_BamHI</b>	GAACAGGATCCCATGTCGTGAACCACTATCCT	Promoteur P <sub>IR</sub> inv

### 3. Expériences de conjugaison

Les bactéries, donatrices et réceptrices, ont été cultivées jusqu'en fin de phase exponentielle (DO 0,8) ou en phase stationnaire (1,6-1,8), puis concentrées après centrifugation à une DO<sub>600</sub> finale de 30, pour enfin être mises en contact sur un filtre de nitrocellulose hydrophobe de 3 mm (Millipore, Billerica) en ratio 1:1. Après une incubation sur la nuit à 37°C, les bactéries ont été resuspendues en milieu TH, diluées puis étalées sur boîte TH, avec les antibiotiques appropriés afin de quantifier le nombre de bactéries donatrices, et le nombre de transconjugants. La fréquence de transfert est calculée en divisant le nombre de transconjugants par le nombre de bactéries donatrices. Lors des expériences de transfert vers d'autres espèces bactériennes, en plus de l'érythromycine, la tétracycline a été utilisée pour sélectionner les transconjugants chez *S. gallolyticus*. *S. agalactiae*, contrairement à *S. macedonicus*, n'est pas capable de se multiplier à 45°C, les transconjugants dans *S. macedonicus* ont donc été sélectionnés sur érythromycine à 45°C.

#### 4. Extraction et analyse bioinformatique des séquences.

Les EGMs apparentés aux *TnGBS* ont été recherchés en utilisant trois critères. (i) Identification de gènes codant pour des DDE transposases, apparentées aux transposases des *TnGBS*, grâce à une analyse par BlastP et tBlastN dans la base de données nr et dans la banque de données de séquences génomiques bactériennes du NCBI. (ii) Identification à proximité de ces gènes, de gènes codant pour des protéines avec des domaines VirB4 et VirD4. (iii) Identification d'IRs apparentées, permettant de déterminer les limites de l'élément, grâce au programme "EMBOSS 6.3.1: palindrome", qui est accessible par internet sur le portail <http://mobylye.pasteur.fr/cgi-bin/portal.py#welcome> et par inspection visuelle. Puis ces éléments ont été classés en sous familles *TnGBS1* ou *TnGBS2*, selon les protéines constituant leur module de conjugaison.

Les séquences des ICEs complets, apparentés aux *TnGBS*, ont été utilisées pour analyser la conservation des protéines de *TnGBS1* et de *TnGBS2* au sein des deux sous-familles, par une approche de BDBH (BiDirectionnel Best Hit), en utilisant comme critère une identité de séquence protéique supérieure à 20% sur au moins 70% de la protéine. Les séquences extraites au NCBI n'étant pas forcément complètement annotées, et certaines protéines des *TnGBS* étant de taille relativement petite (inférieure à 100 AAs), la présence ou non de ces protéines et leur conservation a ensuite été reconfirmée par une analyse par tBlastN, en conservant les mêmes critères que décrits précédemment.

#### 5. Effet de la mitomycine sur la production de la forme circulaire

Les souches NEM316 et les mutants  $\Delta gbs1141$  et  $\Delta gbs1147$  ont été récupérés à  $DO_{600}$  0,4 (phase exponentielle) et à  $DO$  1,4 (phase stationnaire). Après un traitement de 40 minutes à différentes concentrations de mitomycine (0, 20, 50, 100 ng/mL), l'ADN génomique a été extrait, et les formes circulaires de *TnGBS2* ont été quantifiées comme décrit dans le "matériels et méthodes" de l'article n°3.

## 6. **Activité du promoteur P<sub>IR</sub>**

Les promoteurs P<sub>IR</sub> et P<sub>IR\_inv</sub> (Figure 32) ont été clonés dans le vecteur pTCVlac [374] à partir des produits de PCR, qui ont été obtenus avec les oligonucléotides respectifs P<sub>IRT2\_REcoR</sub>-P<sub>IRT2\_FBamH</sub> et P<sub>IRT2inv\_REcoR</sub>-P<sub>IRT2inv\_FBamH</sub>. Ces plasmides ont été introduits dans les souches NEM316 et BM110 et dans le mutant  $\Delta gbs1147$  de la souche NEM316. Les dosages bêta-galactosidase ont été réalisés en utilisant le protocole décrit précédemment [141], en incluant une incubation supplémentaire de 5 min en présence de lysozyme à la concentration finale de 100 $\mu$ g/mL, et un traitement avec du Triton X-100 à 0.1%, comme proposé par Ferrari [375].

## B. **Résultats et discussion**

### 1. **Les TnGBS, une nouvelle famille d'ICE (publication n°3)**

L'objectif dans cette première partie du travail était de démontrer le rôle de la transposase (Gbs1118), et des deux IRs prédites lors de l'étude du génome de la souche NEM316 de *S. agalactiae*, dans le processus de transposition. En 2002, il avait été proposé que les protéines codées par les gènes *gbs0410*(TnGBS1) et *gbs1118* (TnGBS2), qui présentent 48% de similarité au niveau de la séquence protéique, pourraient agir comme des intégrases [125]. Ces protéines possédaient un domaine protéique, présentant un faible taux de similarité avec le domaine rve des intégrases rétrovirales, appartenant à la grande famille des transposases à motif DDE [376]. L'alignement des séquences des protéines Gbs0410 et Gbs1118 avec celles de transposases localisées dans des ISs, apparentées chez d'autres bactéries à Gram positif, a permis de constater la présence d'un motif DDE conservé entre ces protéines, suggérant que ces gènes codent pour une nouvelle famille de transposases à motif DDE. Mathieu Brochet avait montré que l'inactivation du gène *gbs1118* altère la production de l'intermédiaire circulaire de TnGBS2.

J'ai utilisé une stratégie complémentaire, basée sur la construction d'un élément minimum, MiniTnGBS2, contenant la transposase et les deux IRs de TnGBS2 dans un plasmide non répliatif chez *S. agalactiae*, pMutin, possédant un gène de résistance à l'érythromycine (article n°3). Ce plasmide (pMutin::MiniTnGBS2) a une organisation similaire à l'intermédiaire de transposition que constitue la forme circulaire de TnGBS2. Cet élément minimum a été électroporé dans la souche *S. agalactiae* BM110, dépourvue de TnGBS2. L'obtention de colonies



résistantes à l'érythromycine, et ayant intégrées dans leur chromosome pMutin::*MiniTnGBS2*, a permis de démontrer que la transposase et les deux IRs sont nécessaires et suffisantes pour l'insertion par transposition. De plus, l'analyse des séquences de jonction, de part et d'autre du pMutin::*MiniTnGBS2*, indique que l'intégration du mini-transposon, comme l'intégration de *TnGBS2*, conduit à une duplication de 7 à 9 pb au niveau du site d'insertion, en accord avec le mécanisme de transposition. Au sein de ces souches, la production d'une forme circulaire du pMutin::*MiniTnGBS2* a pu être détectée par PCR, et l'analyse de la séquence de cette forme circulaire a montré la présence d'une séquence couplante (séquence localisée entre les deux IRs lors de la formation de l'intermédiaire circulaire), dérivée de la séquence dupliquée observée de part et d'autre du pMutin::*MiniTnGBS2* au site d'insertion. Ces résultats indiquent donc que la transposase et les IRs sont suffisantes pour l'excision de pMutin::*MiniTnGBS2*. Onze sites différents d'insertion de pMutin::*MiniTnGBS2* ont été identifiés à partir de 21 transformants, par séquençage direct sur le chromosome.

L'analyse des sites d'insertion de *TnGBS2* dans les isolats naturels, et après conjugaison, indique que ces ICEs sont insérés dans des régions intergéniques. Cependant, aucune similarité de séquence n'est observée au niveau des séquences dupliquées de part et d'autre de l'élément. L'insertion de l'ICE se fait selon une orientation spécifique, l'IR droite (IR située en aval du gène codant pour la transposase) étant toujours située en amont de l'extrémité 5' du gène adjacent. Une analyse plus précise du lieu d'insertion a permis de constater que l'IR droite était toujours localisée à une distance de 15 à 16 nucléotides, en amont de séquences correspondant potentiellement à des boîtes -35, typiques de régions promotrices, suggérant que ces éléments s'insèrent selon un mécanisme original. Cette spécificité d'insertion est aussi observée avec pMutin::*MiniTnGBS*, indiquant que la transposase et les deux IRs sont nécessaires et suffisantes pour conférer la spécificité d'insertion.

L'insertion au niveau de régions intergéniques est supposée minimiser l'impact lié à l'insertion de l'ICE. Cependant, une telle insertion en amont de séquences promotrices pourrait interférer avec le niveau d'expression du gène en aval. Pour tester cette éventualité, j'ai étudié l'effet de l'insertion de *TnGBS2* et de pMutin::*TnGBS2* sur le niveau d'expression du gène adjacent dans trois transconjugants et trois intégrants par qRT-PCR. Les résultats obtenus suggèrent que l'insertion de *TnGBS2* ou du pMutin::*TnGBS2* n'affecte pas de façon significative l'expression des gènes adjacents.

**Publication n°3**

---

# Atypical association of DDE transposition with conjugation specifies a new family of mobile elements

Mathieu Brochet,<sup>1</sup> Violette Da Cunha,<sup>1</sup>  
Elisabeth Couvé,<sup>1</sup> Christophe Rusniok,<sup>2</sup>  
Patrick Trieu-Cuot<sup>3</sup> and Philippe Glaser<sup>1\*</sup>

<sup>1</sup>Institut Pasteur, Unité de Génétique des Génomes Bactériens, CNRS URA 2171, France.

<sup>2</sup>Institut Pasteur, Unité de Biologie des Bactéries Intracellulaires, CNRS URA 2171, France.

<sup>3</sup>Institut Pasteur, Unité de Biologie des Bactéries Pathogènes à Gram Positif, CNRS URA 2172, 28 Rue du Dr Roux, 75724 Paris cedex 15, France.

## Summary

We describe in *Streptococcus agalactiae* an atypical family of conjugative transposons named TnGBSs which associates DDE transposition and conjugation. We present evidence that the transposition of TnGBS2, the prototype of this family, is catalysed by a new class of DDE transposases that are widespread in Gram-positive bacteria. Remarkably, transposition occurs in intergenic regions, 15 or 16 bp upstream the –35 sequence of promoters, minimizing the burden on the host cell and suggesting an association between transcription and transposition. Transposition catalyses the formation of a circular intermediate that is substrate for subsequent conjugative intercellular transfer. Conjugation is initiated at an origin of transfer by a transposon-encoded relaxase. Whereas all integrative and conjugative elements described so far encode a phage-related integrase, TnGBS2 is the first example of conjugative transposon whose recombination is mediated by a DDE transposase. The combination of DDE transposition with conjugation implies recombination constraints linked to the physical separation of donor and recipient molecules.

## Introduction

Mobile genetic elements are composed of functional modules that allow their maintenance and dissemination, and the regulation of both processes. They also frequently encode adaptive functions that are not directly involved in

their mobility such as antibiotic resistance or alternative metabolic pathways. They play a major role in facilitating lateral gene flow in prokaryotes (Frost *et al.*, 2005) through diverse mechanisms (Salyers *et al.*, 1995) including conjugation (i.e. cell-to-cell contact and transfer of single- or doubled-strand DNA) and recombination (i.e. low or high specificity of integration catalysed by serine or tyrosine recombinases). This diversity has led to a heterogeneous nomenclature, including conjugative transposon, integrative plasmid, genomic island, but numerous elements remain unclassified. This nomenclature is confusing since, for example, the term ‘transposon’ was used for elements whose mobility relies on intrinsically different mechanisms of transposition. To avoid any ambiguity, it has been proposed to group all the elements that are both conjugative and integrative, regardless of specificity of their mechanism, as ‘Integrative Conjugative Elements’ (ICEs) (Burrus *et al.*, 2002a).

Tn916 originally isolated from the Gram-positive bacterium *Enterococcus faecalis* (Franke and Clewell, 1981) is one of the best-characterized ICEs. The first step in Tn916 conjugative transfer is an excision event catalysed by a site-specific tyrosine recombinase (Poyart-Salmeron *et al.*, 1989) that produces a covalently closed circular molecule (Scott *et al.*, 1988). This double-strand circular form could be a substrate for integration in the genome of the donor cell or may be transferred by conjugation to a recipient cell. A nick at the origin of transfer (*oriT*) initiates the transfer of a single-strand intermediate (Jaworski and Clewell, 1995) and transfer is assumed to be replicative. Thus, after completion of the transfer and replication, donor and recipient cells each contain a copy of the double-strand circular molecule (Scott *et al.*, 1994). Finally, the tyrosine recombinase catalyses the integration of the circular form in both cells.

Tn916 was originally designated ‘conjugative transposon’ as it apparently associates conjugative functions with transposition properties similar to those of typical class II transposons, whose mobility is dependent on DDE transposases. However, Tn916 transposition is mediated by a tyrosine site-specific recombinase and the presence of a Ser/Tyr recombinase was observed in all ICEs studied so far. Ser/Tyr integrases and DDE transposases represent mechanistically distinct recombinases. Site-specific tyrosine recombinases form a covalent

Accepted 11 December, 2008. \*For correspondence. E-mail philippe.glaser@pasteur.fr; Tel. (+33) 145 688 996; Fax (+33) 145 688 786.

3'-phosphotyrosine linkage with the DNA leading to the creation of a covalently closed circular molecule. In contrast, the DDE motif represents a catalytic triad of acidic amino acids that ensure DNA cleavage of the terminal phosphodiester bond at the 3' end of the transposon strand and its transfer into the target DNA site (Haren *et al.*, 1999). The DDE transposase family is extremely diverse and transposable prokaryotic DNA elements range from the genetically complex bacteriophage Mu to the elementary insertion sequences (ISs) composed of two short-terminal inverted repeats (IRs) flanking a single or two short open reading frame(s).

*Streptococcus agalactiae*, also referred to as group B *Streptococcus* (GBS), is an opportunistic pathogen that causes severe invasive neonatal infections, such as pneumonia, sepsis and meningitis (Dermer *et al.*, 2004). *S. agalactiae* is also able to colonize mammary glands of various ruminants where it can survive for extended periods (Keefe, 1997). Dissemination and transfer mechanisms of mobile elements within the *S. agalactiae* species remain poorly known. About 80% of human *S. agalactiae* isolates are resistant to tetracycline due to the presence of Tn916 or related ICEs (Poyart *et al.*, 2003). However, the conjugal transfer of Tn916 elements has never been reported between *S. agalactiae* strains under laboratory conditions. In *S. agalactiae*, Tn916 can mobilize a small transposon MTnSag1 encoding lincosamide resistance (Achard and Leclercq, 2007).

Previous comparisons of three complete genome sequences of *S. agalactiae* revealed a composite organization of the chromosome with a stable backbone and 11–14 large chromosomal variable regions (Tettelin *et al.*, 2005; Brochet *et al.*, 2006). Recently, we have shown that most of these variable regions correspond to putative ICEs or related elements (Brochet *et al.*, 2008a). In strain NEM316, genomic islands III, VII and VIII are identical and correspond to a putative ICE designated pNEM316-1 (Glaser *et al.*, 2002). In this strain, island X was shown to mobilize chromosomal DNA through a Hfr-type mechanism, highlighting its potential for conjugative intercellular transfer (Brochet *et al.*, 2008b).

Here, we report that genomic island X and pNEM316-1 constitute a new family of non-composite DDE transposons. We demonstrated that the transposition of these elements is mediated by a novel subfamily of DDE transposases and analysed the insertion specificity of this subfamily into intergenic regions. We also showed that genomic island X is present within a large pool of *S. agalactiae* strains and further characterized its conjugal transfer under laboratory conditions. This unique combination of transposition and conjugation implied specific characteristics of transposition which may enable its coupling with conjugation.

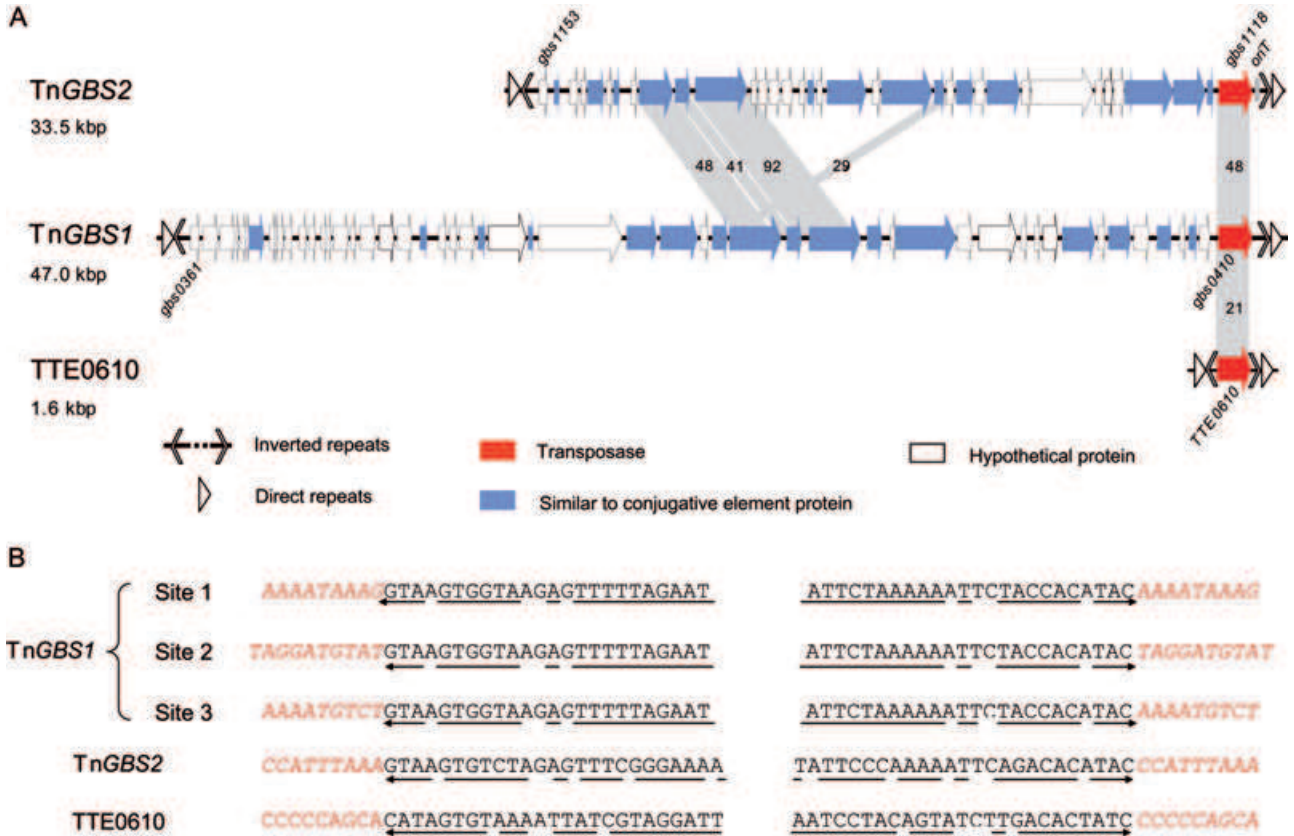
## Results

*In silico analysis suggests that pNEM316-1 (TnGBS1) and genomic island X (TnGBS2) are DDE conjugative transposons*

The three copies of pNEM316-1 present in *S. agalactiae* NEM316 are flanked by 26 bp imperfect IRs and 9 or 10 bp direct repeats (DRs) (Glaser *et al.*, 2002) and encode a protein (Gbs0410) containing a domain weakly similar (Pfam E-value: 0.069) to the rve domain motif of retrovirus integrases that are DDE transposases (Khan *et al.*, 1991). Interestingly, in strain NEM316, genomic island X contains a paralogue of *gbs0410*: *gbs1118* (Fig. 1A). Using the known sequence of the extremities of pNEM316-1, we showed that the genomic island X is delineated by similar IRs (Fig. 1B). In addition, these IRs are directly flanked by a 9 bp repeated sequence (TTTAAATGG), which is a characteristic of DDE transposons. pNEM316-1 is larger (47 kbp) than genomic island X (33.5 kbp) and 50 and 35 putative CDSs were predicted in these elements respectively.

Multiple copies of genes encoding proteins homologous to Gbs0410 or Gbs1118 were detected by blastP searches within the genomes of unrelated Gram-positive bacteria: 23 in *Thermoanaerobacter tengcongensis* MB4, 8 in *Geobacillus kaustophilus* HTA426, 7 in *Halothermotrix orenii* H168, 8 in *Lactobacillus reuteri* 100–23 and 3 in *L. reuteri* F275. These genes are flanked by IRs and DRs. Figure 1A and B show the organization of TTE0610 of *T. tengcongensis* MB4 as an example. Multiple alignments of the deduced amino acid sequences of four of the encoded proteins revealed the conservation of active domain regions with, in particular, the three acidic catalytic residues (aspartyl, aspartyl and glutamyl) of DDE transposases (Fig. 2). These observations suggest that these DDE transposase-encoding genes with the associated IRs represent a new family of ISs widely distributed in Firmicutes. pNEM316-1 and the genomic island X are thus likely transposons and were renamed TnGBS1 and TnGBS2 respectively.

Although the transposases and the IRs of TnGBS1 and TnGBS2 are highly similar, the comparison of both transposons showed that they encode only five homologous proteins (Fig. 1A). In addition to the putative transposases, these proteins include three putative surface exposed LPxTG proteins similar to surface exclusion proteins and aggregation factors described in several conjugative plasmids from Gram-positive bacteria (Hirt *et al.*, 1996). A putative relaxase was detected only in TnGBS2 (Gbs1121). The fact that both transposons encode proteins possessing homologues in unrelated conjugation systems of other Gram-positive bacteria (e.g. five proteins related to Tra determinants of the enterococcal conjugative plasmid pCF10 in TnGBS2) strongly suggests that they might be



**Fig. 1.** General features of TnGBS transposons and of related ISs. A. Genetic organization of TnGBS1, TnGBS2 and of a related IS identified in the *T. tengcongensis* MB4 genome sequence (TTE0610, NC\_003869.1). Arrows indicate the direction of transcription and names of the first and last open reading frames (ORFs) are indicated. Shaded areas indicate significant sequence similarity between ORF products of both transposons. Numbers correspond to the percentage of identity of the corresponding proteins. B. Nucleotide sequences of the extremities of TnGBS transposons and of the IS TTE0610. Black letters and arrows, imperfect inverted repeats; red letters, direct repeats.

conjugative. However, the putative conjugation machineries of TnGBS1 and TnGBS2 are clearly different.

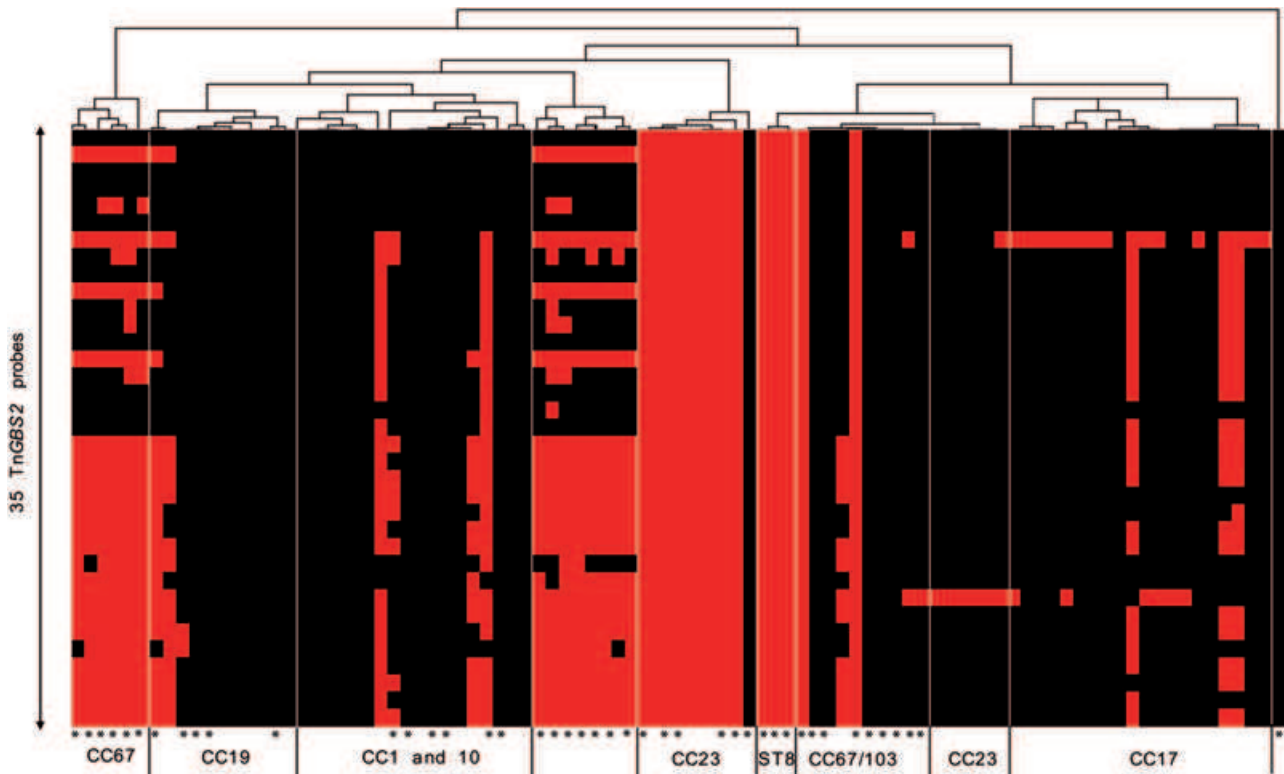
*TnGBSs are widely distributed among S. agalactiae isolates*

TnGBS1 and TnGBS2 were expected to be identified in other *S. agalactiae* isolates since they encode putative

conjugative functions. We thus analysed the presence of both elements in 92 isolates by DNA-array hybridization. These isolates were collected from different origins and belong to all major clonal complexes (CCs) described by multilocus sequence typing (MLST) so far (Brochet *et al.*, 2006; Table S1). We focused our analysis on the 42 and 35 probes specific for TnGBS1 and TnGBS2 respectively. Twelve strains from three different clonal complexes

	┌──────────┐	┌──────────┐	┌──────────┐
Gbs0410	tns <b>D</b> <sub>239</sub> g <b>Gh</b> (61)	tv <b>L</b> <sub>306</sub> <b>D</b> <b>TtEa</b> (33)	fstragigvm <b>E</b> <sub>352</sub> sqhr <b>Ki</b>
Gbs1118	tns <b>D</b> <sub>238</sub> n <b>Gk</b> (61)	tv <b>L</b> <sub>305</sub> <b>D</b> <b>TiEs</b> (33)	isprgigvm <b>E</b> <sub>351</sub> sqhr <b>Kv</b>
TTE0610	vng <b>D</b> <sub>263</sub> - <b>Ga</b> (50)	vf <b>L</b> <sub>318</sub> <b>D</b> <b>TlEa</b> (42)	qklygmgva <b>E</b> <sub>373</sub> ttvd <b>Kk</b>
Lreu0733	las <b>D</b> <sub>243</sub> a <b>Gp</b> (58)	ii <b>L</b> <sub>307</sub> <b>D</b> <b>TyEs</b> (34)	kdihligsv <b>E</b> <sub>354</sub> sshraf

**Fig. 2.** Amino acid sequence alignments of the putative DDE motif of Gbs1118 and related transposases. TTE0610 and Lreu0733 were identified in the *T. tengcongensis* MB4 genome sequence (NC\_003869.1) and in *L. reuteri* 100-23 (NZ\_AAP200000000) respectively. The DDE triad is shown in large bold letters (position of these residues in the protein is indicated) together with the associated downstream conserved residues (Haren *et al.*, 1999). The numbers in parentheses show the number of amino acid residues between the conserved motifs.



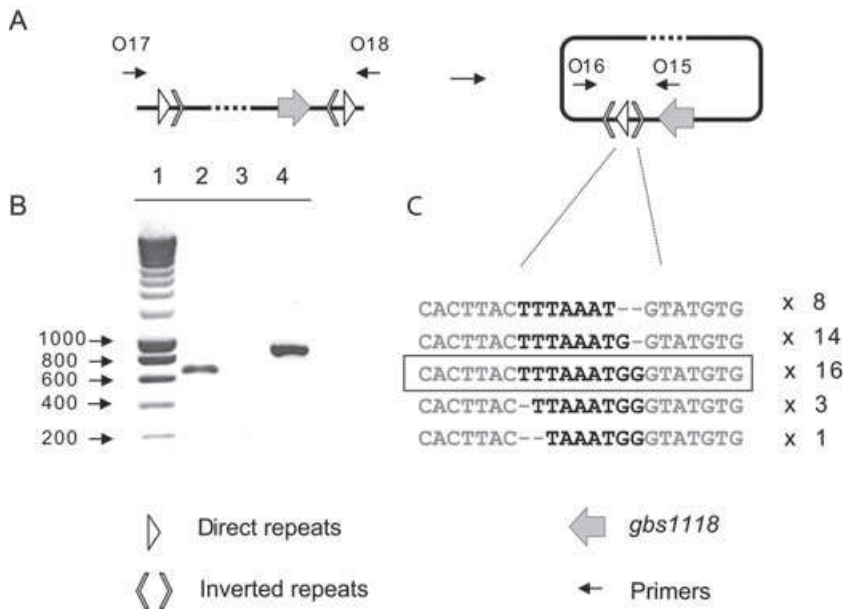
**Fig. 3.** Distribution of the 35 *TnGBS2* ORFs in 92 strains assessed by CGH. Red and black areas denote the predicted presence and absence of genes respectively. The dendrogram indicates the relationships of the 92 strains deduced from their whole gene content by hierarchical clustering (Brochet *et al.*, 2006). Main CCs deduced from MLST data are indicated at the bottom. Stars denote strains isolated from animals.

(CC23, 26 and 67) were positive for all *TnGBS1*-specific probes (Table S1) and 13 strains from CC8, 23 and 67 were positive for the 35 *TnGBS2* probes (Fig. 3 and Table S1) indicating that they probably carry *TnGBS1* or *TnGBS2* respectively. The presence of similar if not identical elements in distantly related CCs suggests their horizontal transfers under natural conditions. We also identified 24 isolates from different CCs showing a heterogeneous pattern of hybridization with 9–27 out of the 35 *TnGBS2* probes being positive. These strains probably carry a mobile genetic element related to *TnGBS2*. For the arbitrarily retained ratio cut-offs, sequences that diverge by more than 20% from the DNA probe were considered as absent (Brochet *et al.*, 2006). It is therefore possible that some genes are too divergent to give a positive signal although homologues are present in the *TnGBS2*-related elements. *TnGBS2* appears as the prototype of a large and diverse family of transposons which are detected in all major CCs. Interestingly, *TnGBS1*, *TnGBS2* and related elements were more frequently detected in strains of bovine origin (44% and 85% respectively) than in human isolates (2% and 16% respectively).

*TnGBS2* was chosen for further detailed analyses in subsequent studies because of its widespread distribution and presence as a unique copy in the genome of the model strain NEM316.

#### *Gbs1118 is necessary and sufficient to mediate the transposition of TnGBS2*

We previously showed that *TnGBS1* is present in NEM316 both as an integrated and a closed circular form (Glaser *et al.*, 2002). By using a similar PCR approach, we tested whether extra-chromosomal circular forms of *TnGBS2* could be detected in the cell (Fig. 4A). Amplification with divergent primers located at both ends of *TnGBS2* (O15 and O16) yields a *c.a.* 500 bp amplicon only if a circular form is present in the whole-cell DNA extract. The expected PCR product was obtained after 30 cycles of amplification (Fig. 4B). Quantification of this circular form by quantitative PCR (Q-PCR) indicated that its relative copy number per chromosome is  $1.25 (\pm 0.27) \times 10^{-4}$ . By direct sequencing of the amplicon with O15 or O16, only the first half of the sequences corresponding to either transposon extremities was readable (data not



**Fig. 4.** Characterization of the TnGBS2 transposition. A. Schematic representation of the linear inserted and of the circular excised forms. Primers O17-O18 were used to detect the absence of an inserted copy, and O15-O16 to detect the circular form by PCR. B. PCR analysis: 1, molecular weight marker in bp (Smart ladder, Eurogentec); 2, O15-O16 (NEM316); 3, O17-O18 (NEM316); 4, O17-O18 (BM110). C. Nucleotide sequence O15-O16 amplicons in 42 clones: grey letters correspond to the extremities of the transposon, black letters to the junction sequence, numbers on the right to the number of clones. The sequence corresponding to the 9 bp chromosomal DR in NEM316 is boxed.

shown). Our inability to read the sequence further was likely due to a heterogeneity of the intervening sequence. In order to characterize the circle junction, the amplicons obtained with primers O15-O16 were cloned into pGEM-T and the sequences of the inserts were determined for 42 clones. Analysis of these sequences revealed that the junction of the circular form comprised both IRs in a convergent orientation, separated by a sequence of variable length (7–9 bp). This intervening sequence derives from the 9 bp duplication (TTTAAATGG) present at the chromosomal junctions of TnGBS2 in NEM316 (Fig. 4C). The length variation in 22 clones is due to the presence of a reduced number of G residues, and to a reduced number of T residues in four clones. These observations suggest that the transposition process may involve strand cleavages at both ends of the transposon or a strand cleavage at one end of the transposon and the attack of the other end by the free 3'OH generated by the first cleavage.

We tested whether excisive circularization of TnGBS2 restores the target DNA by DNA amplification using oligonucleotides located on both sides of the insertion site (O17 and O18, Fig. 4A). In this experiment, the predicted original target site was not amplified with NEM316 total DNA, whereas, in a control experiment, the expected amplicon was obtained when the DNA of BM110, a strain devoid of TnGBS2, was used as template (Fig. 4B). TnGBS2 excision might therefore be replicative, leaving a copy of the transposon embedded in the original target site as observed for IS911 (Duval-Valentin *et al.*, 2004).

To analyse the role of the putative transposase, Gbs1118, in the circularization of TnGBS2, we insertionally inactivated *gbs1118* in the NEM316 chromosome and the production of the circular form of the transposon was

assayed by PCR in the resulting strain GMP202. The circular junction of the transposon was never detected in the mutant strain in repeated experiments (Fig. 4B). This result demonstrated that Gbs1118 is required for production of the extra-chromosomal circular form of TnGBS2.

To determine if Gbs1118 together with TnGBS2 IRs were sufficient to mediate transposition, we cloned a 1764 bp fragment carrying both IRs and *gbs1118* into the pMutin suicide vector (Vagner *et al.*, 1998). The resulting construction pMutin::MiniTnGBS2 represents a circular intermediate of a minimal transposon conferring resistance to erythromycin in *S. agalactiae*. Plasmids pMutin and pMutin::MiniTnGBS2 were electroporated into the strain BM110 that is devoid of sequences related to TnGBS1 or TnGBS2 (Brochet *et al.*, 2006). Transformants resistant to erythromycin were only obtained with pMutin::MiniTnGBS2. In the 21 tested transformants, an integrated copy of pMutin::MiniTnGBS2 was identified by direct sequencing on genomic DNA using oligonucleotides located on both sides of the IRs (O19 and O20). A single copy of the recombinant vector was inserted at 11 different chromosomal sites (Fig. 5). Analysis of the junction sequence of pMutin::MiniTnGBS2 with the chromosome indicated that the integration of the mini-transposon led to a 7–9 bp duplication at the insertion site, as observed for TnGBS2. Gbs1118 was also sufficient to promote circularization in transformants as a pMutin::MiniTnGBS2 circular form was PCR-detected by using primers O23 and O24 which are located on both sides of the junction of the circular form (data not shown). We were able to recover the circular form of pMutin::MiniTnGBS2 by transforming *Escherichia coli* cells with total DNA from the BM110 integrants, which confirms the circularization

	IR-R	DR	-35	-10	NI	TC	MiniTn
<b>TnGBS2</b>							
<i>I</i> <sub>gbs0053</sub>	TATTC <del>CA</del> ATAATTCAGACACATAC	AAATTTGTGAAACTAT	<del>TTACAA</del> AGTTTTTAAATATGC	TATAATAATCT			1
<i>I</i> <sub>gbs0094</sub>	TATTC <del>CA</del> ATAATTCAGACACATAC	TTTTGTGTTTTTTCC	<del>TTGACT</del> ATCTTTTAAAAAGGAG	TATAATAAAA			1
<i>I</i> <sub>gbs0107</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	AAGTATCATTAGCTCTTTTCTTTTCTAAAAATAAGG		TATAATAGGTT	1	2	
<i>I</i> <sub>gbs0209</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	AAATAAAAAGAAATAA	<del>TTGACG</del> AATGCTACCTTATTTT	TGATATGATAT	1		
<i>I</i> <sub>gbs0411</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	CAAAATAAAGAAAGATGA	<del>TTGACT</del> AAAGCCTATAAAACAAG	TAAAATAA			1
<i>I</i> <sub>gbs0413</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	ATTTAAAGTAAATGA	<del>TTGCAA</del> TAATCGTATTTT	TGCTATAATAGTAT	1	2	
<i>I</i> <sub>gbs0453</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	TTTTTCAAAAAGTTG	<del>TTGACA</del> CTTAACGAAATTAG	TATAATAATAT	6	54	5
<i>I</i> <sub>gbs0540</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	GAATAGCGATATAAAA	<del>ATAGCA</del> TAATCGCTATTTT	TGATATAATATAG	1		
<i>I</i> <sub>gbs0694</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	AAATTAGTAAATGTC	<del>TTGAAA</del> TTATAAAAAATAT	TGTATAATAAAAA	1		
<i>I</i> <sub>gbs0782</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	AAAAGCATAAAATAC	<del>TTGCTT</del> TTTTAGAGGATTT	TGTATAATAAGATA	1		
<i>I</i> <sub>gbs1117</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	CCATTTAAATTAATTC	<del>TTGAAA</del> TCTTAATTAAAAAAT	TACAAATAGTT	18	2	
<i>I</i> <sub>gbs1262</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	TTATAACGCAAAAGGC	<del>TTGCAA</del> TTTTTCAGAATATAG	TATAATATATA	2	1	
<i>I</i> <sub>gbs1510</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	AAAATAATCTTCTCAT	<del>TTAGTA</del> TTAATTGATATTATTTT	TATTTTT	1		
<i>I</i> <sub>gbs1579</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	TAATGGCTTCAAAC	<del>TTGCTT</del> TTATTTATGATTATAG	TATAATGGTA	5	3	
<i>I</i> <sub>gbs1584</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	CTAATCTGACAAGAAA	<del>TTGTCG</del> GATTAGTTTTTT	TGTATAATGAAG			1
<i>I</i> <sub>gbs1946</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	ATATATAAGTAAAGGG	<del>TTTACA</del> AAAAACAATAAG	TGCTATAATAAAAC	1	1	
<i>I</i> <sub>IRNA<sup>Lys</sup></sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	AAATCGATAATTTT	<del>TTGAAA</del> ATTTTATAAAAAAG	TGGTATCATAAAG			3
<i>I</i> <sub>san0423</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	ATTTAAAGTAAATGA	<del>TTGCAA</del> TAATCGTATTTT	TGCTATAATAGTAT	1		
<i>I</i> <sub>san2316</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	TAGCTAACAAAATCC	<del>TTTACA</del> TTTAGAACGTGCG	TGATATAATATAGT	1		
<i>I</i> <sub>san0854</sub>	TATTC <del>CA</del> AAAAATTCAGACACATAC	AAAAGCATAAAATAC	<del>TTGCTT</del> TTTTAGAGGATTT	TGTATAATAAGAT	1		
<b>TnGBS1</b>							
<i>I</i> <sub>gbs0411</sub>	TATTC <del>TA</del> AAAAACTCTTACCCTTAC	AAATAAAGAAGATGA	<del>TTGACT</del> AAAGCCTATAAAACAAG	TAAAATGAAGA			
<i>I</i> <sub>gbs0741</sub>	TATTC <del>TA</del> AAAAACTCTTACCCTTAC	TAGGATGTATTGAAAGT	<del>TTCCAA</del> TAACTAGGAAATC	TGTATAATTA			
<i>I</i> <sub>gbs0968</sub>	TATTC <del>TA</del> AAAAACTCTTACCCTTAC	AAATGCTATTATAA	<del>TAGACA</del> TTTTTATTTTGT	TGCTATAATAGAC			

**Fig. 5.** DNA sequences of the insertion sites of TnGBS2, TnGBS2-like transposons, TnGBS1 and pMutin::MiniTnGBS2. *gbs* and *san* correspond to the nomenclature of the orthologous gene in strains NEM316 or COH1 respectively. For TnGBS2 and related elements, numbers of identified insertions per site are indicated on the right for 27 natural isolates (NI), 71 transconjugants (TC) and 21 pMutin::MiniTnGBS2 transformants (MiniTn). Blue letters represent the right extremity of the transposable element (IR-R); Underlined letters, the sequence duplicated after insertion (DR); Red letters, putative -35 and -10 elements; purple letters, putative extended -10 region.

of the inserted copy. In these plasmids, the sequence of the circular junction was different from that of the original pMutin::MiniTnGBS2 plasmid but corresponded to the duplicated DR in the corresponding BM110 integrant (data not shown). These results demonstrated that Gbs1118 and its associated IRs are sufficient to mediate TnGBS2 transposition.

#### TnGBS2 is self-transmissible by conjugation

To assess the conjugal transfer of TnGBS2, we inserted the pG<sup>+</sup>host5 plasmid encoding erythromycin resistance upstream *gbs1118* in NEM316. The recombinant strain was designated GMP201. Conjugation experiments were conducted by mating the GMP201 donor strain with BM110 used as a tetracycline-resistant recipient strain. A transfer frequency of  $4.6 \times 10^{-6}$  was observed when

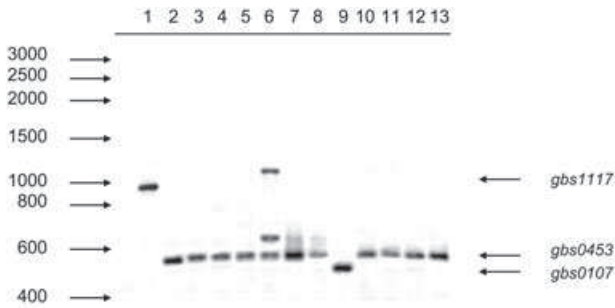
transconjugants were selected with erythromycin and tetracycline (Table 1). Given the proportion of cells with a circular form of the transposon ( $1.25 \times 10^{-4}$ ), this value indicates a high efficiency of transfer of the circular intermediate. Five independent transconjugants were characterized by DNA-hybridization to the NEM316

**Table 1.** Transfer efficiencies of recombinant TnGBS2 according to different donor and recipient strains.

Donor strain	Recipient	Transfer efficiency
GMP201 NEM316-TnGBS2::erm	BM110	$4.6 \pm 0.4 \times 10^{-6}$
		$4.9 \pm 0.3 \times 10^{-6a}$
BM110-TnGBS2::erm	A909RF	$9.8 \pm 0.2 \times 10^{-7}$
GMP202 NEM316Δ <i>gbs1118</i> ::erm	BM110	$7.88 \pm 0.2 \times 10^{-11}$
GMP203 NEM316Δ <i>gbs1121</i> ::erm	BM110	$1.07 \pm 0.1 \times 10^{-10}$

a. In the presence of deoxyribonuclease I.





**Fig. 6.** Southern blot hybridization of 13 *Sau3A*-digested chromosomal DNA with the O21-O22 probe. Lane 1: NEM316, a single band of *c.a.* 800 bp corresponding to the insertion at  $I_{gbs1117}$  is visible while the circular form is not detected due to its low copy number per cell; Lane 2, 3, 4, 5, 7, 8, 10, 11, 12 and 13: BM110 transconjugants in which a single integration occurred at  $I_{gbs0453}$ ; Lane 6, BM110 transconjugant in which three integrations occurred; Lane 9: BM110 transconjugant in which a single integration occurred at  $I_{gbs0107}$ .

whole-genome DNA-array. The hybridization pattern of transconjugants was similar to that of strain BM110 with a positive signal for the 35 probes corresponding to *TnGBS2* (data not shown). These results demonstrated that these transconjugants corresponded to strain BM110 carrying an entire copy of *TnGBS2*. Furthermore, Southern blot analysis of the transconjugants indicated that *TnGBS2* integrated into the chromosome after its conjugal transfer (Fig. 6).

We investigated whether the *TnGBS2* inserted into the chromosome of BM110 was able to retransfer to a secondary recipient strain. Transfer of the erythromycin marker was observed at a frequency of  $9.8 \times 10^{-7}$  when BM110 transconjugants were filter-mated with the recipient strain A909RF (Table 1). DNA-array hybridization confirmed that transconjugants corresponded to strain A909RF carrying a whole copy of *TnGBS2* (data not shown). These results suggest that *TnGBS2* carries all the functions necessary for its horizontal transmission. In addition, these transfers were not sensitive to the presence of DNase I in the mating medium, which excludes the possibility that a transformation-like mechanism is involved (Table 1).

#### Identification of functions required for the transfer of *TnGBS2*

To investigate the role of Gbs1118 in intercellular transfer, we conducted matings using GMP202 as donor strain. This strain is deficient for *TnGBS2* transposase activity and unable to produce the circularized element. The transposase deficiency led to a 60 000-fold reduction of transfer efficiency leading to a basal level of transfer frequency (Table 1) which corresponds to chromosomal mobilization, as previously described (Brochet *et al.*,

2008b). This result suggested that the transposase activity is required for conjugal transfer of *TnGBS2*.

The relaxase is a central component of most conjugation systems which initiates transfer by cleaving one *oriT* DNA strand. Within *TnGBS2*, *gbs1121* was predicted to encode a 543 aa protein similar to plasmid relaxases and was shown to be involved in the conjugative transfer of DNA regions located downstream *TnGBS2* (Brochet *et al.*, 2008b). To determine whether Gbs1121 is required for *TnGBS2* transfer, the corresponding gene in NEM316 was insertionally inactivated and we investigated whether the resulting strain GMP203 was still able to transfer the transposon. A dramatic reduction in *TnGBS2* transfer frequency was observed when GMP203 was mated with BM110, although few transconjugants were obtained (Table 1). As for the transposase mutant, this basal level of transfer frequency was shown to correspond to chromosomal mobilization (data not shown). Quantification of the amount of the circular form relative to the integrated copy by Q-PCR showed that the circularization of *TnGBS2* was not affected in GMP203 ( $1.03 \pm 0.06$  fold change compared with the wild type). These results indicate that Gbs1121 is required for *TnGBS2* transfer without interfering with the formation of the circular form.

Sequences of *oriT* regions are not conserved, but they often contain direct and inverted repeats and are typically located in intergenic regions close to the DNA processing genes (Grohmann *et al.*, 2003). Among the 10 intergenic regions of *TnGBS2* longer than 100 bp, we identified a 308 bp sequence containing several direct and inverted repeats. This sequence is located between the IR-R and the stop codon of *gbs1118*. To test whether this region (*oriT<sub>TnGBS2</sub>*) could initiate a conjugal transfer, we cloned it together with the two IRs into plasmid pTCV-*erm* (Poyart and Trieu-Cuot, 1997) that is non-mobilizable in NEM316. Matings were performed using NEM316 carrying either pTCV-*erm*Δ*oriT<sub>TnGBS2</sub>* or pTCV-*erm* as donors and BM110 as a recipient. Efficient transfers of pTCV-*erm*Δ*oriT<sub>TnGBS2</sub>* were detected while pTCV-*erm* was not mobilized (Table 2). To confirm that the putative *oriT<sub>TnGBS2</sub>* was recognized by Gbs1121, we tested whether pTCV-*erm*Δ*oriT<sub>TnGBS2</sub>* was mobilizable in strain GMP204 carrying an in-frame deletion of the relaxase gene (*gbs1121*). We

**Table 2.** Mobilization of pTCV-*erm* or pTCV::*oriT<sub>TnGBS2</sub>* to strain BM110.

Donor strain	Plasmid	Transfer efficiency
NEM316	pTCV- <i>erm</i>	ND <sup>a</sup> ( $< 5.2 \pm 0.4 \times 10^{-11}$ )
NEM316	pTCV- <i>erm</i> Δ <i>oriT<sub>TnGBS2</sub></i>	$2.1 \pm 0.3 \times 10^{-8}$
GMP204 NEM316	pTCV- <i>erm</i> Δ <i>oriT<sub>TnGBS2</sub></i>	ND ( $< 3.7 \pm 0.2 \times 10^{-11}$ )
	Δ <i>gbs1121</i> :: <i>erm</i>	

a. No transconjugant detected.

did not detect any transfer of pTCV-*erm*Ω $\text{oriT}_{\text{TnGbs2}}$  from GMP204::pTCV-*erm*Ω $\text{oriT}_{\text{TnGbs2}}$  to BM110 (Table 2). Together, these results strongly suggest that Gbs1121 is required to initiate conjugal transfer from *oriT* $\text{TnGbs2}$ .

#### *TnGbs2* integrates upstream putative –35 promoter sequences

The integration site of *TnGbs2* was investigated in 73 BM110 transconjugants by direct sequencing of genomic DNA (Fig. 5). A single insertion was observed in 71 transconjugants. Fifty-three insertions (73%) occurred in a preferential site different from the insertion site of *TnGbs2* in NEM316 and located 488 bp upstream *gbs0453*. This insertion site was thus designated  $I_{\text{gbs0453}}$ . In 18 transconjugants, *TnGbs2* inserted in 12 other sites scattered along the chromosome. For the two remaining strains, double and triple insertions were observed by Southern blot hybridization (Fig. 6, data not shown).

We then further investigated the insertion specificity of *TnGbs2* by direct sequencing of genomic DNA of the 12 natural isolates identified to carry *TnGbs2* by hybridization. In nine strains, *TnGbs2* was inserted at the same position as in NEM316, 204 bp upstream *gbs1117* initiation codon ( $I_{\text{gbs1117}}$ ). In two strains, the insertion site was identified 488 bp upstream the orthologue of *gbs0453* ( $I_{\text{gbs0453}}$ ). In strain CCH350, both sites were occupied by a copy of *TnGbs2*. The insertion site of *TnGbs2*-related elements was also determined in 15 strains (Table S1). Interestingly, four and eight elements were also inserted at  $I_{\text{gbs0453}}$  and  $I_{\text{gbs1117}}$  respectively. This suggests that these transposons probably encode a transposase related to Gbs1118 although the signal of the *gbs1118* probe was negative. Three additional insertion sites were identified in the three remaining strains (Fig. 5).

All identified insertion sites of *TnGbs2* and of pMutin::Mini*TnGbs2* were located in intergenic regions longer than 100 bp. Comparison of the insertions sites did not reveal any significant sequence similarity among the duplicated sequences (Fig. 5). This observation suggested that a mechanism other than sequence-specific integration must be responsible for this targeting into intergenic regions. We noticed that integration of *TnGbs2* always occurred in a specific orientation in which the right extremity of the transposon is located upstream the 5' end of the adjacent gene. Whereas no conserved feature was identified upstream the IR-L extremity, typical –35 and –10 regions of sigma A promoters were predicted downstream the IR-R sequence. The transposon was inserted at a fixed position 15 or 16 bp upstream of the putative –35 elements. *TnGbs2*-related elements and *TnGbs1* were similarly inserted into intergenic regions with similar orientation and position specificities (Fig. 5). The analysis of the chromosomal sequence adjacent to the IR-R of the

**Table 3.** Effect of the integration of *TnGbs2* or pMutin::Mini*TnGbs2* on the transcription of the downstream gene in strain BM110.

Transposable element	Insertion site	Fold change compared with BM110	Standard deviation <sup>a</sup>
<i>TnGbs2</i>	$I_{\text{gbs0453}}$	1.56	1.82
	$I_{\text{gbs1117}}$	–1.88	0.20
	$I_{\text{gbs1579}}$	1.01	0.01
pMutin::Mini <i>TnGbs2</i>	$I_{\text{gbs0053}}$	1.74	1.14
	$I_{\text{gbs1262}}$	2.26	1.05
	$I_{\text{gbs1579}}$	–1.52	0.02

a. Standard deviations correspond to two biological replicates.

ISs related to *TnGbs2* showed in most cases a similar targeting 15 or 16 bp upstream –35 promoter sequences (Fig. S1). This suggests that the specific targeting is an intrinsic property of this new family of transposases.

Insertion into intergenic regions is expected to minimize the mutational effects caused by the transposon insertion. However, such insertions may interfere with the transcription level of the downstream gene. We tested the effect of insertion on the transcription level of the downstream gene in three *TnGbs2* transconjugants and in three pMUTIN-Mini*TnGbs2* integrants. Comparison of the mRNA concentrations of the downstream gene by Q-PCR with the BM110 parental strain revealed that insertion of *TnGbs2* or of pMUTIN-Mini*TnGbs2* did not affect significantly its transcription level (Table 3).

## Discussion

Conjugative transposons or ICEs in Gram-positive and Gram-negative bacteria combine recombination (i.e. the two flanking attachment sites and the phage-related integrase) and conjugation modules. The transfer of known ICEs includes (i) the excision of the integrated copy to form a non-replicative covalently closed circular element, (ii) a nick of the circular intermediate at an *oriT*, (iii) the replicative transfer of a single-strand DNA molecule to regenerate a doubled-strand circular element in the donor and recipient cells and (iv) the integration of the non-replicative transposon in both cells. To our knowledge, all ICEs described so far encode a tyrosine-based lambdaoid integrase or a resolvase family recombinase (Churchward, 2002). Here, we described a new family of conjugative transposons in which the enzyme catalysing the circularization and the integration is not a phage-related integrase but a DDE-type transposase. We further characterized the conjugation process of *TnGbs2*, a representative member of this family identified in *S. agalactiae* strain NEM316. The transfer of *TnGbs2* from a donor to a recipient cell involves the generation of a circular form in the donor cell and a relaxase acting on a *cis*-located

*oriT*-like sequence. These features are similar to those described for the transfer of conjugative plasmids or ICEs such as *Tn916*, but the mechanism of the strand exchange process catalysed by the DDE transposase must comply with an intercellular transposition event in which donor and target molecules are in two different cells. Excision and integration are therefore decoupled implying a physical separation of the transpososomes associating the transposase with the DNA of donor and recipient replicons, successively.

Sequence analysis of the transposase encoded by *TnGBS2* revealed the presence of the three DDE catalytic residues essential for the transposition process. We identified genes encoding similar transposases in uncharacterized ISs from various Gram-positive bacteria. Transposases of this new subfamily do not exhibit significant similarity to known DDE transposases. No other protein motif such as a helix-turn-helix motif was predicted, highlighting the uniqueness of this subfamily. As described for most ICEs, a circular form is produced in the donor cell, which serves as a substrate for the conjugation machinery. In *TnGBS2*, the transposase *Gbs1118* is necessary and sufficient to generate the circular form of the transposon as it was not detected in a *gbs1118* mutant but was generated by a mini-transposon expressing only this protein. Conversely, the circular form is probably a substrate for the transposition in the recipient cell as the electroporated mini-transposon corresponding to a circular transposable element is efficiently inserted into the chromosome.

The mechanism of strand cleavage and transfer mediated by DDE transposases is typified by *IS911*, a member of the *IS3* family, whose transposition involves circularization and replication of one transposon DNA strand and leaves an integrated copy of the IS in the donor molecule (Duval-Valentin *et al.*, 2004). This model is fully compatible with the data obtained for *TnGBS2*, including the characterization of a circular coupling sequence with variable end points at both transposon extremities and our inability to PCR amplify a resealed donor molecule generated during transposition. However, we cannot rule out the possibility that *TnGBS2* transposition proceeds through a cut-and-paste mechanism (e.g. *IS10* transposition mechanism) generating a double-strand circular transposon and a potentially lethal double-strand break in the donor chromosome. A replicative model of transposition involving single-strand cleavages at both transposon ends to generate an element covalently attached to donor and target replicons is not compatible with the conjugative transfer of the transposon, as the donor and the target replicons are in two different cells.

The target selectivity of mobile elements encoding a DDE transposase is greatly variable depending upon the element considered, ranging from high (e.g. *Tn7*) to low, but was never reported to be random (Craig, 1997).

Analysis of integration sites in natural isolates and transconjugants of *S. agalactiae* indicates that *TnGBS2* inserts exclusively in intergenic regions. Similar insertion behaviour was observed for the pMUTIN-Mini*TnGBS2*, *TnGBS1* and different ISs encoding a transposase related to *Gbs1118*. Although these target sites did not display any significant sequence homology, we noticed that the IR-R of the inserted element was always located 15 or 16 bp upstream the  $-35$  region of the promoter of chromosomal genes. Preferential insertion into active promoter regions has been previously described for the retrotransposon *Ty3* in *Saccharomyces cerevisiae* which inserts immediately upstream of genes transcribed by polymerase III (Chalker and Sandmeyer, 1992). *In vitro* experiments have demonstrated that this targeting involves protein-protein interactions between the transposase and the transcription factor TFIIB or TFIIC (Kirchner *et al.*, 1995). Our observations suggest a similar targeting mechanism involving direct interaction of the transposase and the RNA polymerase machinery during initiation. The  $-35$  and  $-10$  promoter regions are recognized by the RNA polymerase holoenzyme containing sigma 70. The region upstream the  $-35$  region is more accessible and has been defined as the UP element (Browning and Busby, 2004). UP elements are found in some promoters and increase their activity by providing a point of contact for the C-terminal domain of the RNA polymerase alpha subunit (Ross *et al.*, 2001). Sequences of UP elements are not conserved but are AT rich. Most of the promoter targeted by *TnGBS2* are canonical, show an extended  $-10$  element (a TG motif located upstream of the  $-10$  region), and have an AT-rich sequence upstream of the  $-35$  region. As *TnGBS2* and related transposable elements are inserted right upstream putative UP regions, it is tempting to speculate that this region is involved in the target specificity and that the alpha subunit of the RNA polymerase interacts with the transposase. As a similar targeting was observed for all ISs and transposons of this subfamily, this specific targeting and the hypothesized interaction with the RNA polymerase are probably intrinsic properties of this subfamily of transposases.

Analysis of the insertion sites of *TnGBS2* and *TnGBS2*-like elements in natural isolates indicated one preferential chromosomal insertion site located upstream *gbs1117*. Interestingly, after conjugation under laboratory conditions into strain BM110, 73% of the transconjugants were inserted upstream *gbs0453*. In the circular element, the junction sequence corresponding to  $I_{gbs1117}$  is therefore not a determining factor in the target specificity of the transposon.

Transposases of *TnGBSs* and related ISs show sequence and size similarities which suggest that their transposition mechanisms are related. It is worth noting that the stop codon of the IS transposase gene is adjacent

**Table 4.** Bacterial strains and plasmids used for cloning and conjugation assays.

Strains and plasmids	Relevant genotype and resistance profile <sup>a</sup> or characteristics	Reference or source
<i>E. coli</i> DH5 $\alpha$		Invitrogen
<i>S. agalactiae</i> NEM316	Isolated from neonate blood culture	Glaser <i>et al.</i> (2002)
GMP201	Erm; NEM316-Tn <i>GBS2</i> ::erm	This study
GMP202	Erm; NEM316 <i>gbs1118</i> ::erm	This study
GMP203	Erm; NEM316 <i>gbs1121</i> ::erm	This study
GMP204	NEM316 <i>gbs1121</i>	This study
BM110	Tet; Isolated from neonate blood culture	Stalhammar-Carlemalm <i>et al.</i> (1993)
A909RF	A909 spontaneous mutant for Rif and Fus	Brochet <i>et al.</i> (2008b)
Plasmids		
pGEM-T	Amp	Promega
pMutin	Erm; Amp	Vagner <i>et al.</i> (1998)
pTCV-erm	Erm, Km; <i>oriR</i> pACYC184, <i>oriRpAM</i> $\beta$ 1	Poyart and Trieu-Cuot (1997)
pGhost5+	Erm; ColE1 replicon, thermosensitive derivative of pGK12	Biswas <i>et al.</i> (1993)

a. Amp, Erm, Fus, Rif, Tet, are resistance to ampicillin, erythromycin, fusidic acid, rifampicin and tetracycline respectively.

to IR-R whereas it is separated by a 310 and 308 bp region in Tn*GBS1* and Tn*GBS2* respectively. In Tn*GBS2*, and probably in Tn*GBS1*, this intervening segment contains the *oriT* sequence that allows conjugative transfer.

Comparison of Tn*GBS1* and Tn*GBS2* showed a conserved recombination module, i.e. a DDE transposase and IRs, but distinct conjugation modules suggesting a modular evolution. By comparative genomic hybridization we identified numerous elements that share putative modules with Tn*GBS2* suggesting that the two Tn*GBS* transposons identified in strain NEM316 are part of a diverse family as described for other ICEs (Roberts *et al.*, 2001; Burrus *et al.*, 2002b; 2006). Characterization of these related elements will reveal the diversity of module combinations within this new family, allowing the study of the interaction between different modules. DNA-array hybridization of a set of diverse strains from both human and animal origins revealed that Tn*GBS* transposons are frequently present within the *S. agalactiae* population. Interestingly, these transposons were more frequently found among strains of bovine or goat origin. The high frequency of Tn*GBS* transposons in strains of animal origin might be explained by a reservoir specific to the rumen of polygastric mammals. Alternatively, these transposons may provide a selective advantage in this environment. Tn*GBS*s were identified among all major CCs. The presence of Tn*GBS2* in a cluster of related strains of different origins within CC23 strongly suggests a vertical transmission of the transposon indicative of a stable insertion in natural isolates. Conversely, its presence in other unrelated strains at different insertion sites probably results from horizontal dissemination under natural conditions. These observations suggest both horizontal transfer and vertical transmission of Tn*GBS2*.

In conclusion, we describe a new subfamily of transposases which show two properties identified for the first

time in prokaryotes. First, these DDE transposases are associated with conjugation modules to form conjugative transposons. This combination raises new issues such as the temporal regulation of both processes and the evolution of a transposition mechanism to cope with a physical separation of the donor and the recipient molecules. Second, these transposases target precisely DNA sequences upstream sigma A promoters which probably minimizes the burden on the bacterial host associated with transposition. Our understanding of transcription initiation in bacteria provides the rational bases for the *in vitro* reconstitution of transposition, as well as for the characterization of the protein–protein interactions and the molecular mechanisms involved in the coupling of transposition and transcription initiation.

## Experimental procedures

### Bacterial strains and growth conditions

*Streptococcus agalactiae* strains used to characterize Tn*GBS2* are listed in Table 4. The strains used to assess Tn*GBS*s distribution are described in Table S1. *E. coli* DH5 $\alpha$  (Invitrogen, Paisley, UK) was used for cloning experiments. *S. agalactiae* was grown in Todd-Hewitt (TH) broth or agar (Difco Laboratories, Detroit, MI) and *E. coli* in Luria–Bertani medium. Antibiotics were used at the following concentrations: for *E. coli* – ampicillin, 100  $\mu$ g ml<sup>-1</sup>; erythromycin, 150  $\mu$ g ml<sup>-1</sup>; for *S. agalactiae* – erythromycin, 10  $\mu$ g ml<sup>-1</sup>; rifampicin, 100  $\mu$ g ml<sup>-1</sup>; fusidic acid, 10  $\mu$ g ml<sup>-1</sup>; tetracycline, 10  $\mu$ g ml<sup>-1</sup>. When specified, 10 units of deoxyribonuclease I (Roche, Basel, Switzerland) were added during mating.

### General DNA techniques

Plasmid DNA was prepared by alkaline lysis using the Qiagen Miniprep plasmid kit and genomic DNA was extracted using the Qiagen DNeasy kit (Qiagen, Hilden, Germany). The oligonucleotide primers used for PCR reactions were designed

with Primer3 software (<http://primer3.sourceforge.net>). To determine insertion sites of TnGBSs, ISs and derived recombinant transposons, sequencing reactions were directly performed on genomic DNA. DNA labelling, DNA-array and Southern blot hybridization and detection were performed as previously described (Brochet *et al.*, 2006).

### Construction of bacterial strains

The primers used for plasmid constructions are listed in Table S2. In order to select TnGBS2 transfers, we tagged the transposon by single cross-over insertion of the pG<sup>+</sup>host5 plasmid encoding resistance to erythromycin (Biswas *et al.*, 1993). For this purpose, the EcoRI-BamH1 O1-O2 amplicon encompassing the 5' terminal part of *gbs1118* was cloned into pG<sup>+</sup>host5 and the recombinant plasmid was introduced by electroporation into NEM316. Transformants were selected on erythromycin agar plates at 30°C. Cells in which the recombinant vector had integrated into the chromosome were selected by growth at 40°C in the presence of erythromycin.

Insertional inactivation of *gbs1118* and *gbs1121* was carried out by single cross-over. In-frame deletion in *gbs1121* was constructed by using splicing-by-overlap-extension PCR, as previously described (Dramsi *et al.*, 2006). EcoRI-BamH1 amplicons were cloned into the thermosensitive shuttle plasmid pG<sup>+</sup>host5. Transformants and integrants were selected as described above. For the *gbs1121* in frame deletion, integrant strains were serially passaged for 5 days at 30°C without erythromycin selection to allow the excision of the plasmid. Dilutions of the serially passaged cultures were tested for erythromycin sensitivity to identify clones in which pG<sup>+</sup>host5Δ*gbs1121* excised. All insertional inactivations and the in-frame deletion were confirmed by sequencing.

For the pMutin::MiniTnGBS2 construction, a fragment containing both IRs and *gbs1118* was amplified from TnGBS2 circular form. The EcoRI-BamH1 amplicon was cloned into the pMutin suicide vector (Vagner *et al.*, 1998). The fragment containing the putative origin of transfer was amplified using O13 and O14 oligonucleotide primers and was cloned into the replicative shuttle vector pTCV-*erm* (Poyart and Trieu-Cuot, 1997). To determine the different junction sequences of TnGBS2 circular form, the O13-O14 amplicon was cloned into the pGEM-T vector.

### Bacterial conjugations

The donor and the recipient strains were mixed together upon a 3 mm hydrophobic edge membrane filter (Millipore, Billerica, MA) with a donor-to-recipient ratio of 10:1. Plates were incubated overnight at 37°C. The bacterial cells were resuspended in TH broth and spread on TH agar plates containing appropriate antibiotics. To quantify conjugative transfer efficiency, dilutions were plated on TH media containing antibiotics for the selection of transposon-containing donor cells. The frequency of transfer was determined by dividing the number of TnGBS transconjugants by the number of donor cells. All conjugation assays were performed at least twice. The gene content of transconjugants was analysed by DNA-array hybridization when appropriate (Brochet *et al.*, 2006).

### Quantitative PCR

Circular form of TnGBS2 was quantified by comparing the amount of circular TnGBS2 DNA (primers O23-O24) with the amount of circular and integrated TnGBS2 DNA (primers O25-O26). The relative quantity of specific DNA was normalized by the relative amplification of O23-O24 and O25-O26 using pMutin::MiniTnGBS2 as a template in which these two regions are in equal amounts. DNA fragments were amplified with the ABI PRISM 7900 SDS system and SYBR Green PCR kits (Applied Biosystems, Foster City, USA). All measurements were performed at least in duplicate. Modification of the expression level of genes located downstream TnGBS2 or pMutin::MiniTnGBS2 was determined by comparing the expression levels in the BM110 transconjugants or transformants with the BM110 parental strain. mRNAs were extracted in mid-exponential growth phase. Q-PCR was performed with *gbs1262* (O29-O30), *gbs0453* (O31-O32), *gbs1579* (O33-O34), *gbs1117* (O35-O36), *gbs0053* (O37-O38) and *gyrA* (O28-O29) specific primers (Table S2). The quantity of specific cDNA was normalized to the quantity of *gyrA* cDNA.

### Acknowledgements

We wish to thank Mike Chandler, Isabelle Rosinski-Chupin, Didier Mazel, Sophie Payot and Gérard Guédon for their interest and critical comments and Carmen Buchrieser for critical reading of the manuscript. We thank Suey van Baarle for her initial contribution to the characterization of TnGBS2. This work received financial support from the Institut Pasteur (GPH 9). M.B. is supported by a grant from the CNRSI (France).

### References

- Achard, A., and Leclercq, R. (2007) Characterization of a small mobilizable transposon, MTn*Sag1*, in *Streptococcus agalactiae*. *J Bacteriol* **189**: 4328–4331.
- Biswas, I., Gruss, A., Ehrlich, S.D., and Maguin, E. (1993) High-efficiency gene inactivation and replacement system for gram-positive bacteria. *J Bacteriol* **175**: 3628–3635.
- Brochet, M., Couvé, E., Zouine, M., Vallaey, T., Rusniok, C., Lamy, M.C., *et al.* (2006) Genomic diversity and evolution within the species *Streptococcus agalactiae*. *Microbes Infect* **8**: 1227–1243.
- Brochet, M., Couvé, E., Glaser, P., Guedon, G., and Payot, S. (2008a) Integrative conjugative elements and related elements are major contributors to the genome diversity of *Streptococcus Agalactiae*. *J Bacteriol* **190**: 6913–6917.
- Brochet, M., Rusniok, C., Couve, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., *et al.* (2008b) Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proc Natl Acad Sci USA* **105**: I15961–I15966.
- Browning, D.F., and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**: 57–65.
- Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002a) Conjugative transposons: the tip of the iceberg. *Mol Microbiol* **46**: 601–610.
- Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002b)

- The ICESt1 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.
- Burrus, V., Marrero, J., and Waldor, M.K. (2006) The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid* **55**: 173–183.
- Chalker, D.L., and Sandmeyer, S.B. (1992) Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev* **6**: 117–128.
- Churchward, G. (2002) Conjugative transposons and related mobile elements. In *Mobile DNA II*. Craig, N., Craigie, R., Gellert, M., and Lambowitz, A. (eds). Washington, DC: American Society for Microbiology Press, pp. 177–191.
- Craig, N.L. (1997) Target site selection in transposition. *Annu Rev Biochem* **66**: 437–474.
- Dermer, P., Lee, C., Eggert, J., and Few, B. (2004) A history of neonatal group B *Streptococcus* with its related morbidity and mortality rates in the United States. *J Pediatr Nurs* **19**: 357–363.
- Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prevost, M.C., Kojadinovic, M., *et al.* (2006) Assembly and role of pili in group B streptococci. *Mol Microbiol* **60**: 1401–1413.
- Duval-Valentin, G., Marty-Cointin, B., and Chandler, M. (2004) Requirement of IS911 replication before integration defines a new bacterial transposition pathway. *EMBO J* **23**: 3897–3906.
- Franke, A.E., and Clewell, D.B. (1981) Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of 'conjugal' transfer in the absence of a conjugative plasmid. *J Bacteriol* **145**: 494–502.
- Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**: 722–732.
- Glaser, P., Rusniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Msadek, T., *et al.* (2002) Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol Microbiol* **45**: 1499–1513.
- Grohmann, E., Muth, G., and Espinosa, M. (2003) Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev* **67**: 277–301.
- Haren, L., Ton-Hoang, B., and Chandler, M. (1999) Integrating DNA: transposases and retroviral integrases. *Annu Rev Microbiol* **53**: 245–281.
- Hirt, H., Wirth, R., and Muscholl, A. (1996) Comparative analysis of 18 sex pheromone plasmids from *Enterococcus faecalis*: detection of a new insertion element on pPD1 and implications for the evolution of this plasmid family. *Mol Gen Genet* **252**: 640–647.
- Jaworski, D.D., and Clewell, D.B. (1995) A functional origin of transfer (*oriT*) on the conjugative transposon Tn916. *J Bacteriol* **177**: 6644–6651.
- Keefe, G.P. (1997) *Streptococcus agalactiae* mastitis: a review. *Can Vet J* **38**: 429–437.
- Khan, E., Mack, J.P., Katz, R.A., Kulkosky, J., and Skalka, A.M. (1991) Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res* **19**: 851–860.
- Kirchner, J., Connolly, C.M., and Sandmeyer, S.B. (1995) Requirement of RNA polymerase III transcription factors for *in vitro* position-specific integration of a retrovirus-like element. *Science* **267**: 1488–1491.
- Poyart, C., and Trieu-Cuot, P. (1997) A broad-host-range mobilizable shuttle vector for the construction of transcriptional fusions to beta-galactosidase in gram-positive bacteria. *FEMS Microbiol Lett* **156**: 193–198.
- Poyart, C., Jardy, L., Quesne, G., Berche, P., and Trieu-Cuot, P. (2003) Genetic basis of antibiotic resistance in *Streptococcus agalactiae* strains isolated in a French hospital. *Antimicrob Agents Chemother* **47**: 794–797.
- Poyart-Salmeron, C., Trieu-Cuot, P., Carlier, C., and Courvalin, P. (1989) Molecular characterization of two proteins involved in the excision of the conjugative transposon Tn1545: homologies with other site-specific recombinases. *EMBO J* **8**: 2425–2433.
- Roberts, A.P., Johanesen, P.A., Lyras, D., Mullany, P., and Rood, J.I. (2001) Comparison of Tn5397 from *Clostridium difficile*, Tn916 from *Enterococcus faecalis* and the CW459tet (M) element from *Clostridium perfringens* shows that they have similar conjugation regions but different insertion and excision modules. *Microbiology* **147**: 1243–1251.
- Ross, W., Ernst, A., and Gourse, R.L. (2001) Fine structure of *E. coli* RNA polymerase–promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev* **15**: 491–506.
- Salyers, A.A., Shoemaker, N.B., Stevens, A.M., and Li, L.Y. (1995) Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* **59**: 579–590.
- Scott, J.R., Kirchman, P.A., and Caparon, M.G. (1988) An intermediate in transposition of the conjugative transposon Tn916. *Proc Natl Acad Sci USA* **85**: 4809–4813.
- Scott, J.R., Bringel, F., Marra, D., Van Alstine, G., and Rudy, C.K. (1994) Conjugative transposition of Tn916: preferred targets and evidence for conjugative transfer of a single strand and for a double-stranded circular intermediate. *Mol Microbiol* **11**: 1099–1108.
- Stalhammar-Carlemalm, M., Stenberg, L., and Lindahl, G. (1993) Protein Rib: a novel group B streptococcal cell surface protein that confers protective immunity and is expressed by most strains causing invasive infections. *J Exp Med* **177**: 1593–1603.
- Tettelin, H., Massignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'. *Proc Natl Acad Sci USA* **102**: 13950–13955.
- Vagner, V., Dervyn, E., and Ehrlich, S.D. (1998) A vector for systematic gene inactivation in *Bacillus subtilis*. *Microbiology* **144**: 3097–3104.

## Supporting information

Additional supporting information may be found in the online version of this article.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## **2. Analyse comparative de la diversité des ICEs de la famille TnGBS (publication n°4 et 5).**

En 2008, Mathieu Brochet n'avait pas identifié d'ICEs homologues à TnGBS1 et TnGBS2 dans les bases de données. Il avait néanmoins montré que des ICEs apparentés à TnGBS1 ou à TnGBS2 étaient fréquents dans les souches de *S. agalactiae* d'origine bovine. Cette analyse, réalisée en collaboration avec Romain Guérrillot, repose sur le fait qu'avec l'augmentation du nombre de séquences génomiques disponibles au NCBI, j'ai pu identifier de nombreux ICEs apparentés aux TnGBS. Les EGMs étant organisés en modules fonctionnels [237], l'analyse de la diversité des ICEs apparentés à TnGBS et des plasmides proches, nous a permis de mieux définir les protéines constituant les modules de transposition, de mobilisation et de réplication. Ces analyses ont permis de mieux comprendre l'évolution de cette famille d'ICEs et de discuter le rôle de chacune de ses fonctions dans sa dispersion, son maintien et sa persistance au sein de son hôte.

### **a) TnGBS1 et TnGBS2 ont une spécificité d'hôte restreinte**

Nous savions que TnGBS2 pouvait être transféré efficacement par conjugaison au sein de *S. agalactiae* (article n°3), cependant nous ne savions pas si c'était également le cas de TnGBS1. Pour tester cette hypothèse, j'ai tout d'abord obtenu une souche de *S. agalactiae* possédant un TnGBS1 en copie unique, et dans lequel un marqueur érythromycine était inséré dans une région intergénique (GMP205). Puis j'ai montré que TnGBS1 pouvait être transféré efficacement, par conjugaison de la souche GMP205, vers la souche réceptrice BM110. Ces deux éléments étant conjuguatifs, ils peuvent servir de modèles pour l'étude de la famille d'ICEs.

L'identification d'éléments apparentés aux TnGBS chez d'autres espèces bactériennes a permis de constater que des ICEs de cette famille sont répandus au sein du genre streptocoque. En particulier, j'ai observé des éléments très similaires dans des espèces différentes. Ce résultat suggère que ces éléments ont été capables de se transférer entre différentes espèces de streptocoques au cours de leur évolution. J'ai pu confirmer expérimentalement dans les conditions du laboratoire le transfert de TnGBS2, avec néanmoins une efficacité réduite vers *S. gallolyticus*, et *S. macedonicus*. Par contre, je n'ai pas réussi à obtenir de transferts de TnGBS2

vers *L. lactis* et *E. faecalis*. TnGBS1 peut être transféré vers *S. gallolyticus* avec une efficacité très faible, mais n'est pas transmis à *L. lactis* et *E. faecalis*. Ces résultats montrent que ces ICEs ont une spécificité d'hôte restreinte, ce type de spécificité restreinte étant également observé pour les plasmides pAD1 et pCF10 d' *E. faecalis* [316, 317].

## b) **Base de données de séquences d'ICEs de type TnGBS**

La première étape de ce projet a été l'établissement d'une base de données de séquences d'ICEs, apparentés aux TnGBS. J'ai d'abord combiné des recherches par BlastP et tBlastN en utilisant la transposase de TnGBS2 et les protéines de conjugaison de TnGBS1 et TnGBS2, pour rechercher ces ICEs dans les génomes disponibles au NCBI. J'ai ainsi pu identifier 56 éléments complets ou partiels apparentés à TnGBS1 ou 2. J'ai également identifié, dans quatre génomes de *S. agalactiae* séquencés au laboratoire, des ICEs de cette famille. J'ai ensuite pu identifier, dans la majorité des cas, les IRs ainsi que le site d'insertion et donc définir les limites de l'ICE. L'analyse des séquences IR montre une grande conservation, en particulier de l'IR gauche, avec néanmoins une différenciation entre les IRs des ICEs apparentés à TnGBS1 ou à TnGBS2. L'analyse des sites d'insertion montre que ces éléments ont la même spécificité d'insertion en amont de promoteurs dépendant de  $\sigma A$ . Finalement, afin de comparer les fonctions codées par ces ICEs, j'ai récupéré les séquences protéiques pour les génomes annotés, et j'ai annoté les ICEs qui ne l'étaient pas encore, essentiellement par similarité avec TnGBS1 et TnGBS2.

## c) **Conservation protéique et analyse phylogénétique**

L'analyse des éléments apparentés aux TnGBS suggère qu'ils peuvent être classés en deux sous-familles d'ICEs, représentées respectivement par TnGBS1 et TnGBS2, en fonction des protéines constituant leur appareil de transfert. Au sein de ces deux sous-familles, 34 (sur 49) et 14 (sur 36) protéines sont respectivement conservées par l'ensemble des membres de la sous-famille TnGBS1 et TnGBS2. Ces résultats démontrent également une plus grande diversité des membres de la sous-famille TnGBS2, ce qui suggère que TnGBS1 serait apparu plus récemment



que *TnGBS2*. Les gènes codant pour les protéines les plus conservées, au sein des différents membres de la famille *TnGBS2*, sont les gènes codant pour les protéines avec des domaines *VirB4*, *VirD4*, la relaxase, et la transposase. De façon surprenante, un gène codant pour une protéine avec un domaine N-terminal *repA* est également retrouvé chez tous les éléments apparentés à *TnGBS2*, suggérant l'importance de la réplication au cours du processus de conjugaison. Les deux seules protéines conservées entre les membres de la sous-famille *TnGBS2* et qui sont absentes des plasmides apparentés, sont les protéines codant la transposase et *Gbs1146*, de fonction inconnue. Cette observation suggère que la protéine codée par *gbs1146*, qui ne présente aucune similitude avec une autre protéine connue, ne serait pas impliquée dans le processus de conjugaison ou de réplication. De plus, elle n'est pas nécessaire à la transposition de l'ICE (article n°3), il serait alors important d'analyser le rôle de cette protéine. Nous avons également constaté que les gènes potentiellement impliqués dans le processus de régulation sont beaucoup moins conservés au sein des éléments de la famille des *TnGBS2*, ce qui indique que ces éléments utilisent différentes stratégies pour la régulation de leur transfert. De façon similaire pour *TnGBS1*, on constate que sont conservés, par l'ensemble des éléments de la sous-famille, les gènes codant pour les protéines de l'appareil de transfert, comme les protéines avec les domaines *virB4*, *virB6* et *virD4*, les gènes codant pour la transposase et la relaxase, et également des gènes codant pour des protéines impliquées dans le processus de réplication, avec des domaines *ParA* ou *RepE*. Ces observations nous ont conduit à proposer que *TnGBS1* et *TnGBS2* étaient probablement capables de répliquions au cours du processus de conjugaison.

L'analyse de la conservation des protéines de *TnGBS2* suggère également des phénomènes de recombinaisons, entre éléments apparentés à *TnGBS2*, au cours de l'évolution de ces éléments. Cette observation a été confirmée par les analyses phylogénétiques réalisées par Romain Guérillot avec les séquences protéiques de la transposase, de la relaxase et des protéines avec des domaines *VirB4* et *VirD4*. L'arbre des transposases confirme que l'ensemble des transposases récupérées d'ICEs sont regroupées au sein du même groupe phylogénétique, suggérant une origine commune à ces transposons. Cet arbre montre également que les transposases, issues d'ICEs apparentés à *TnGBS1*, sont regroupées au sein d'un même sous-groupe phylogénétique, ce qui indique que *TnGBS1* aurait émergé à partir d'un élément *TnGBS2*, qui aurait échangé son module de transfert avec un autre EGM, par exemple un plasmide ou un autre ICE.

Ces différentes informations nous ont conduit à regrouper les gènes de ces ICEs en quatre modules fonctionnels différents impliqués dans la transposition, la mobilisation, le transfert (système de sécrétion de type 4), la réplication et la régulation. Nous avons donc par la suite cherché à analyser chacun de ces modules et Romain Guérillot à valider expérimentalement certains d'entre eux.

#### **d) La réplication de *TnGBS1* et de *TnGBS2* dans la bactérie réceptrice participe au retransfert de l'ICE.**

Différentes observations nous ont permis de montrer que *TnGBS1* et *TnGBS2* sont capables de se maintenir dans la bactérie réceptrice de façon transitoire sous forme circulaire (article n°5). (i) Des séquences couplantes, provenant du site d'insertion de l'ICE lors de l'excision de celui-ci dans la bactérie donatrice, sont retrouvées chez des transconjugants de *TnGBS1* et de *TnGBS2*, ce qui confirme l'absence d'insertion de l'ICE dans ces transconjugants après transfert. (ii) La quantification des formes circulaires de *TnGBS2* révèle que ces formes sont plus abondantes dans les transconjugants que dans la bactérie donatrice. Afin de tester si le maintien sous forme circulaire pouvait jouer un rôle dans la propagation de l'ICE, j'ai étudié le transfert de *TnGBS2* en utilisant des transconjugants comme bactéries donatrices. Les résultats obtenus montrent une augmentation de l'efficacité de transfert, qui peut être corrélée avec l'augmentation du nombre de formes circulaires dans la cellule donatrice. Ces observations confirment le rôle limitant de la production de l'intermédiaire circulaire pour le transfert de *TnGBS2*. Les mêmes expériences réalisées par Romain Guérillot avec *TnGSB1* montrent que *TnGBS1* se réplique également dans la cellule réceptrice. L'analyse du phénomène d'intégration, au cours de passages successifs sur boîtes, montre que l'insertion de *TnGBS2* est beaucoup plus rapide que celle de *TnGBS1*.

Ces observations fonctionnelles sont en accord avec l'identification de protéines de réplication de type RepA, à la fois dans *TnGBS1* et *TnGBS2*, et de type RepE, dans *TnGBS1*. Ces modules de réplication sont donc fonctionnels dans certaines conditions.

En complément de ces analyses, Romain Guérillot a, identifié les *oriT* à la fois de *TnGBS1* et *TnGBS2*, par similarité avec les séquences *oriT* connues des plasmides possédant des

modules de conjugaisons apparentés à ceux des ICEs de type Tn*GBS1* et Tn*GBS2*. Il a également montré que le clonage de ces *oriT* dans un plasmide non mobilisable, dérivé de pAM $\beta$ 1, permettait la mobilisation vers des souches de *S. agalatae*, mais également vers une souche d' *E. faecalis*. Aucun transfert de Tn*GBS1* ou de Tn*GBS2* n'est observé dans *E. faecalis*. Le transfert du plasmide pTCVerm contenant l'*oriT* de Tn*GBS1* ou de Tn*GBS2* est lui obtenu à haute fréquence. Ce qui indique que l'absence de transfert de l'ICE est, dans ce cas, probablement liée à un défaut de maintien dans la bactérie réceptrice, et non pas à une incompatibilité de l'appareil de transfert. Ceci illustre l'existence d'une seconde étape clé pour la conjugaison et la propagation de l'ICE, c'est-à-dire son maintien, par réplication ou par intégration dans la bactérie réceptrice.

En conclusion, cette étude montre que les ICEs apparentés aux Tn*GBS* se sont largement répandus au sein des espèces du genre streptocoque. Cependant leur nombre est probablement encore sous-estimé, et leur propagation au sein du genre streptocoque l'est probablement également. Par exemple, bien qu'aucun Tn*GBS* n'ait été encore identifié chez *S. mutans*, l'analyse du locus CRISPR révèle la présence d'espaceurs dirigés contre Tn*GBS1*, suggérant que cette espèce a été en contact avec des éléments apparentés aux Tn*GBS* [377]. La présence d'un élément de cette famille dans une souche de lactobacille confirme que notre vision de la diversité des ICEs de cette famille est sans doute partielle. L'expansion de cette famille suggère que ces éléments apportent un avantage sélectif, cependant ces éléments ne codent pas pour des gènes de résistance aux antibiotiques. Parmi les protéines qui pourraient conférer un avantage sélectif à la bactérie receveuse, il faut noter la présence d'une adhésine, annotée comme facteur d'agrégation, et appartenant à une famille dont un membre est impliqué dans l'attachement de la bactérie à la dent, la formation de biofilms, et la modulation de l'immunité chez *S. mutans* [256].

## **Publications n°4 et 5:**

---

## CHAPTER

---

# Integrative and Conjugative Elements Encoding DDE Transposases

Violette Da Cunha,<sup>1</sup> Romain Guérillot,<sup>1</sup> Mathieu Brochet<sup>2</sup>  
and Philippe Glaser<sup>\*1</sup>

### Abstract

Conjugative transposons were known for many years and were initially discovered for their capacity to disseminate antibiotic resistance genes. They were first named transposons even if they rely on phage-type integrases for their mobility and are now classified as Integrative and Conjugative Elements (ICE). Recently, two families of mobile genetic elements combining DDE transposition and conjugation have been described for the first time in *Streptococcus agalactiae* (TnGBSs) and in *Staphylococcus aureus* (ICE6013), respectively. These two families of ICEs are widespread in their respective host species but show a restricted host range. The conjugation imposes constraints on the transposition mechanisms as donor and recipient DNA molecules are in two different cellular compartments. It requires in particular the formation of an independent circular intermediate of transposition. Analysis of TnGBS transfer revealed that conjugation mechanisms are similar to that of integrase encoding ICEs despite unrelated excision and insertion mechanisms.

### Introduction

Until recently, the mobility of all characterized Integrative and Conjugative Elements (ICEs) was shown to rely on a conserved mechanism involving an integrase (also called recombinase) and, in some cases an excisionase.<sup>1</sup> ICE excision occurs in a donor cell and leads to a covalently closed circular molecule. A relaxase subsequently initiates transfer by nicking duplex DNA at a specific *oriT* sequence and by transferring a single stranded DNA molecule to the recipient cell in association with the type 4 secretion system (T4SS) of the conjugation machinery. Transferred DNA is afterward converted to a double strand recircularized DNA molecule by the cellular replication machinery. Insertion of the circular ICE is finally catalyzed by the integrase in the donor and the recipient cells.<sup>1</sup> In ICEs, genes involved in a same function like recombination or conjugation are grouped in distinct functional modules. Each mobile element corresponds to a particular combination of such modules and combinatorial shuffling between modules allows for the assembly of highly diverse elements. ICEs share properties with plasmids and phages, and were initially named with a heterogeneous nomenclature including conjugative transposon or integrative plasmid.<sup>2</sup> They may use a serine recombinase, like Tn5397 from *Clostridium difficile*<sup>3</sup> or a tyrosine recombinase like Tn916 that is widespread among Gram-positive bacteria.<sup>4</sup> Despite a diverse range of module combinations, no ICE was reported to depend on a DDE type transposase for its mobility. Although transposases and integrases catalyze co-integration of two

<sup>1</sup>Laboratoire Evolution et génomique Bactériennes, CNRS URA2171, Institut Pasteur, Paris, France; <sup>2</sup>The Wellcome Trust Sanger Institute, Cambridge, UK.

\*Corresponding Author: Philippe Glaser—Email: pglaser@pasteur.fr

DNA molecules, the mechanisms of action are intrinsically different. In contrast to site-specific recombinases, transposases do not covalently bind to DNA.

Recently, ICEs relying on a transposase for their mobility have been described in the human opportunistic pathogens *Streptococcus agalactiae* and *Staphylococcus aureus*. Two ICEs, Tn*GBS1* and Tn*GBS2* are integrated in the genome of *S. agalactiae* strain NEM316.<sup>5</sup> Both ICEs express a closely related DDE transposase belonging to a new class but encode two different conjugation machineries. In *S. aureus* ICE6013 expresses an IS30-like transposase and carries a conjugative module similar to that of ICE*Bs1*.<sup>6</sup> In the present chapter we discuss the specificity of these two families of conjugative elements, the putative constraints of the conjugation process explaining the paucity of reported cases of such an association and the diversity of these elements together with their relationship to other conjugative elements.

### Organization and Diversity of Tn*GBS1* and Related Transposons

Tn*GBS1* is 47 kb long and carries 49 protein encoding genes. All genes except one (*gbs0376\**) are transcribed in the same orientation. It is present in three copies in *S. agalactiae* strain NEM316. Its GC content of 37.6% is rather uniform along the sequence and is slightly higher than the average value of the NEM316 genome (35.6%). The transposase gene is located at the right end of the element with the stop codon located 335 bases far from the end of the transposon (Fig. 1). Among these 49 genes, 16 encode proteins similar to proteins encoded by diverse conjugative plasmids and putative ICEs identified in various Firmicutes. In addition to the transposition module (the transposase and the two inverted repeats—IRs), different modules were identified (Fig. 1). The mobilization module was defined by similarity with the predicted relaxase and the *oriT* sequence from pHT $\beta$  and pMG2200 plasmids.<sup>7</sup> It includes the gene encoding the putative relaxase (*gbs0384*) preceded by four genes encoding proteins of unknown function. An *oriT* site similar to the *oriT* sequence characterized in pHT $\beta$  was identified in the promoter region of *gbs0380* (Fig. 1). This mobilization module represents a new type not previously categorized.<sup>8</sup> In addition, two genes encoding a putative topoisomerase I (*gbs0387*) and a protein containing a primase domain (*gbs0386*) could also be involved in the DNA processing during intercellular transfer. The conjugation module was identified by the presence of a VirD4 homolog (*gbs0396*), a distantly related VirB4 protein (*gbs0402*), and genes encoding surface proteins homologous to aggregation and exclusion proteins from Gram-positive conjugative plasmids.<sup>9</sup> In addition, a protein with a predicted CHAP domain (Gbs0404) is possibly implicated in the cleavage of the peptidoglycan for building type IV structures across the wall of the bacteria, as previously reported for the CsiA protein encoded by *Lactococcus lactis* MG 1363 sex factor.<sup>10</sup> Tn*GBS1* encodes also two proteins homologous to plasmid replication proteins: RepA from diverse Gram-positive plasmids and a pAM $\beta$ 1 RepE homolog<sup>11</sup> (Fig. 1). The prediction of these functional modules suggests that Tn*GBS1* is conjugative and may under certain circumstances be able to replicate autonomously. Accordingly, Tn*GBS1* was shown to be transferable by conjugation as Tn*GBS1* carrying an inserted erythromycin resistance genes (*erm*) was readily transferred among *S. agalactiae* strains with a frequency of 10<sup>-5</sup>–10<sup>-6</sup> per donor cell (V. Da Cunha and R. Guérillot, unpublished).

DNA-array hybridizations of a collection of 92 *S. agalactiae* isolates of human and animal origins revealed the presence of this transposon or of a related transposon in 12 strains. These strains were all of animal origin but belong to distant clonal complexes (CC23, 26 and 67).<sup>5</sup> Until recently, there was no sequence similar to Tn*GBS1* in public sequence libraries. However, with the increasing number of genomes available, search for similar sequences revealed the presence of seven complete or partial sequences of ICEs similar to Tn*GBS1* (Table 1). Six of them were found in streptococci and a more distantly related one in *Lactobacillus salivarius*. In this last case, genes similar to Tn*GBS1* genes are located on two contigs (Table 1). These transposons share the functions predicted in Tn*GBS1* including transposition, mobilization, conjugation and replication.

\*For Tn*GBS1* gene names correspond to the first of the three copies of this ICE in the genome of strain NEM316

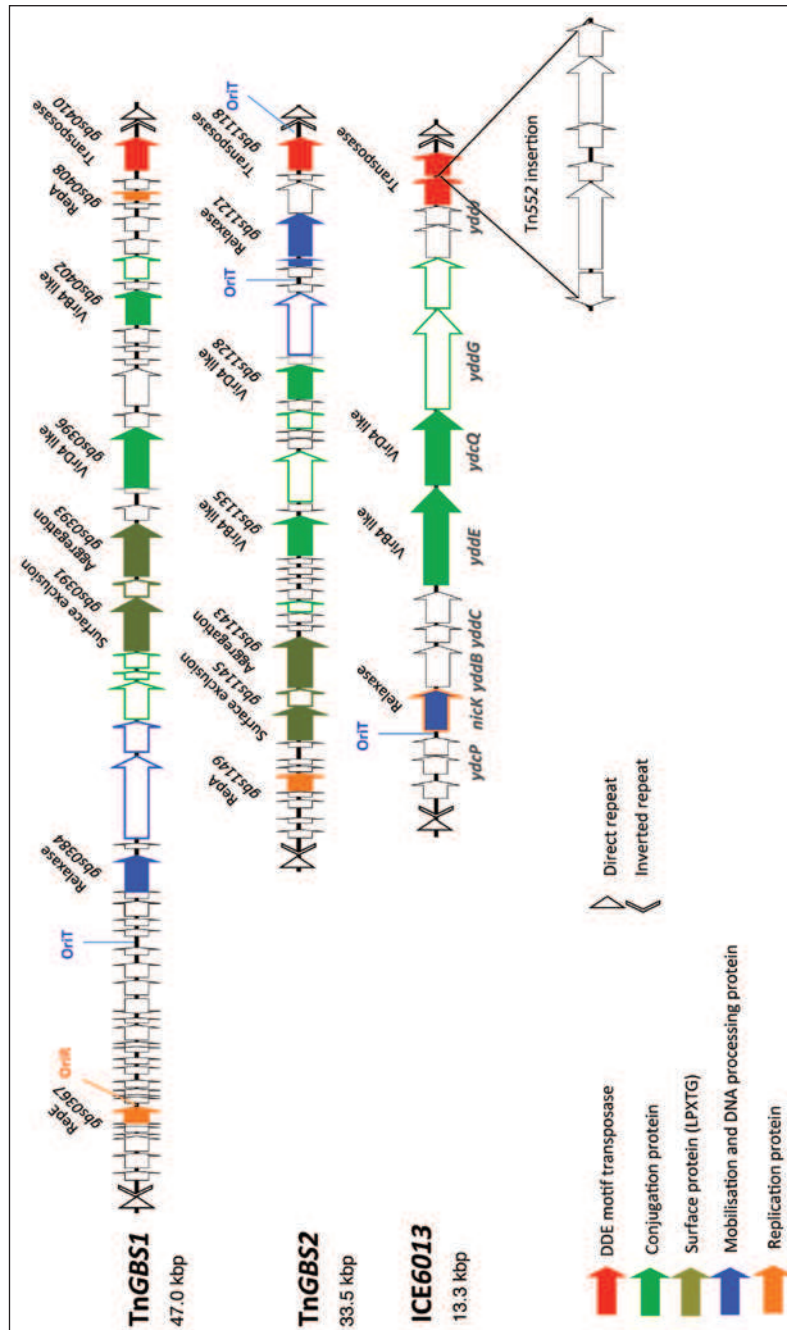


Figure 1. General organization of TnGBS1, TnGBS2 and ICE6013. The functional modules are defined according to the color code shown on the figure. Colored filled arrows correspond to genes with predicted functions. Colored empty arrows indicate genes predicted to belong to a functional module without defined function. Black arrows represent genes of unknown functions. TnGBS1 is present in three copies in strain NEM316, gbs numbers are given for the first copy. For ICE6013, the gene names underneath the arrows correspond to the names of ICEBs1 orthologous genes.

Table 1. General characteristics of ICEs related to *TnGBS1* and *TnGBS2*

Species—Strain	Type	Insertion Site <sup>1</sup>	Size kb (nb of genes)	Cargo Gene	Acc. Number
<i>S. agalactiae</i> NEM316	<i>TnGBS1</i>	<i>gbs0411-gbs0741-gbs0968</i>	47 (49)		NC_004368
<i>S. agalactiae</i> NEM316	<i>TnGBS2</i>	<i>xpt</i>	33 (36)		NC_004368
<i>S. agalactiae</i> ATCC 13813	<i>TnGBS1</i>	<i>galE</i>	partial		AEQQ01000019
<i>S. agalactiae</i> ATCC 13813	<i>TnGBS2</i>	<i>glyQ</i>	36 (31)		AEQQ01000097–98
<i>S. infantarius</i> ATCC BAA-102	<i>TnGBS2</i>	<i>rpmH</i>	64 (70)	Heavy metal resistance	NZ_ABJK02000018
<i>S. gallolyticus</i> UCN34	<i>TnGBS2</i>	<i>rpmB</i>	43 (45)	Sortase A	NC_013798
<i>S. gallolyticus</i> ATCC BAA-2069-A	<i>TnGBS2</i>	<i>rpmH</i>	36 (37)	Glucan-binding protein	NC_015215
<i>S. gallolyticus</i> ATCC BAA-2069-B	<i>TnGBS2</i>	ND <sup>2</sup>	30 (35)		NC_015215
<i>S. oralis</i> ATCC 35037	<i>TnGBS2</i>	ND	29 (30)		NZ_ADMV01000022
<i>S. taxon oral sp taxon 071 str. 73H25AP</i>	<i>TnGBS1</i>	<i>rpsO</i>	46 (44)		NZ_AEEP01000006
<i>S. taxon oral sp taxon 071 str. 73H25AP</i>	<i>TnGBS2</i>	<i>glyQ</i>	31 (31)		NZ_AEEP01000011
<i>Streptococcus sp 2_1_36FAA</i>	<i>TnGBS1</i>	<i>glmS</i>	40 (41)		NZ_GG704942
<i>Streptococcus sp 2_1_36FAA</i>	<i>TnGBS2</i>	<i>rpsF</i>	25 (27) partial		NZ_GG704942
<i>S. mitis</i> SK597	<i>TnGBS2</i>	<i>pyrF</i>	32 (32)		NZ_AEDV01000027
<i>S. sanguinis</i> ATCC 49296	<i>TnGBS2</i>	<i>rpmB</i>	32 (30)		NZ_AEPO01000010

continued on next page



Table 1. Continued

Species—Strain	Type	Insertion Site <sup>1</sup>	Size kb (nb of genes)	Cargo Gene	Acc. Number
<i>S. sanguinis</i> SK1058	TnCBS1	<i>pyrC</i>	Partial		AFBF01000017
<i>S. sanguinis</i> SK49	TnCBS1	<i>relA2</i>	Partial		AFFO01000015
<i>S. anginosus</i> F0211	TnCBS1	<i>murF</i>	50(50)		AECT01000063
<i>S. ictaluri</i> 707-05	TnCBS2	<i>gbs1262</i> orthologous gene	28 partial		AEU01000023
<i>S. ictaluri</i> 707-05	TnCBS2	<i>rimP</i>	37 partial		AEU01000037
<i>S. ictaluri</i> 707-05	TnCBS1	<i>rodA</i>	45		AEU01000014
<i>S. ictaluri</i> 707-05	TnCBS1	<i>rodA</i>	45		AEU01000004
<i>L. salivarius</i>	TnCBS1	ND	32 + 73 partial	Pilus locus,	NZ_ACGT01000033 NZ_— ACGT01000030

1. Name of the gene upstream which the ICE is inserted. 2. Not determined

They are therefore likely mobile. Additional genes encoding proteins of unknown function are conserved and may be involved in the transfer process. For example, these elements express a highly conserved ClpA homolog. ClpA is the ATPase/protein unfoldase subunit of the ATP-dependent ClpAP protease.<sup>12</sup> The chaperone activity of this protein may be required for DNA transfer by this atypical conjugation machinery.

### Organization and Diversity of Tn*GBS2* and Related Elements

Tn*GBS2* is 33.5 kb long and is predicted to encode 36 genes. Its G+C content of 38% is, like for Tn*GBS1*, slightly higher than the *S. agalactiae* core genome G+C content. As observed in Tn*GBS1*, all genes except *gbs1134* are transcribed in the same orientation (Fig. 1). The transposase gene is located 335 bases from the right-end of the transposon. The mobilization module belongs to the MOB<sub>p7</sub> class<sup>8</sup> and encodes two proteins: a relaxase (Gbs1121) and a relaxase accessory protein (Gbs1122) as defined previously.<sup>5</sup> We also predict a conserved sequence similar to previously characterized *oriT* regions upstream this module<sup>13</sup> (R. Guérillot, unpublished data). This sequence may represent a second *oriT* region of Tn*GBS2* in addition to the one previously identified downstream the transposase gene.<sup>5</sup> Comparative analysis allowed identification of a conjugative module similar to that of the pheromone inducible conjugative plasmid pCF10.<sup>5</sup> Therefore Tn*GBS2* encodes a T4SS unrelated to the one identified in Tn*GBS1*. The conjugation module encodes all essential proteins identified for the conjugative process of pCF10 (Fig. 1 and ref. 14). In agreement with this observation, Tn*GBS2* is efficiently transferred among *S. agalactiae* strains with a frequency of 10<sup>-5</sup> per donor cell.<sup>5</sup> Despite the unrelated T4SS used by Tn*GBS2* and Tn*GBS1*, three proteins carrying a peptidoglycan anchoring motif (LPXTG) and possibly involved in the DNA transfer are highly similar in both ICEs: Gbs1143 is 92% identical to Gbs0393 (annotated as a putative cell aggregation protein), Gbs1144 is 41% identical to Gbs0392 and Gbs1145 (annotated as a putative exclusion protein) is 48% identical to Gbs0391. This level of similarity reflects the likely exchange of these genes between different conjugation machineries and may be linked to the host specificity of these elements among streptococci. The conjugation process involves the formation of a circular intermediate catalyzed by the transposase as observed in Tn*GBS1*.<sup>5</sup> The abundance of the circular form of Tn*GBS2* was estimated by quantitative PCR as 10<sup>-4</sup> copies per chromosome. This value is in a similar range of the conjugation frequency. Tn*GBS2* is therefore efficiently transferred by conjugation and its circularization represents the limiting step. Like in Tn*GBS1*, we identified a putative *repA* gene suggesting that it may also replicate during the conjugation process.

DNA-array analysis of a collection of strains showed that Tn*GBS2* and related transposons are common among *S. agalactiae* isolates.<sup>5,15</sup> For unknown reason, it was observed as particularly abundant among isolates of animal origin, where it was detected in 85% of the strains. Consistently, only NEM316 carries a Tn*GBS2* element among the eight published genomes of *S. agalactiae*, all of human origin. This puzzling observation may correspond to a counter selection for strains carrying this element among human isolates.<sup>5</sup> The CRISPR-cas system provides immunity against phages and mobile genetic elements.<sup>16</sup> The protection is mediated by short RNA (crRNA) transcribed from short sequences (spacers) captured from the invaders and inserted into the CRISPR locus. Spacers targeting Tn*GBS2* were identified in four published genomes (2603V/R, A909, CJB111, 18RS21, V. Da Cunha, Unpublished). The distribution of Tn*GBS2* among *S. agalactiae* strains may therefore be controlled by the immunity provided by the CRISPR-cas system.

Similarity searches at the time of the initial characterization of Tn*GBS2* failed to detect any similar transposon among sequenced genomes. However, with the increasing number of published genomes, we have now identified 13 Tn*GBS2*-like elements in different streptococcal genomes (Table 1). These transposons are frequent in the pyogenic, bovis and mitis groups (Table 1). The transposition, mobilization and conjugation modules are conserved, and all these elements express a RepA homolog. These Tn*GBS*-related elements are thus likely mobile apart from the ICE identified in *Streptococcus infantarius* in which several of the genes essential for mobility are pseudogenes or are missing.

An important question is the selective advantage these transposons may bring to their host cell. Only the ICE identified in *Streptococcus infantarius* expresses a typical cargo gene annotated as conferring resistance to heavy metals (Table 1), but this ICE is probably non-conjugative. We identified in TnGallo1 from *Streptococcus gallolyticus* strain UCN34 a *srtA* paralogous gene encoding a type A sortase.<sup>17</sup> The sortase A catalyzes the anchoring of surface proteins carrying an LPXTG motif to the peptidoglycan. The function of this second Sortase A protein is unclear, but may contribute to the fitness of the host by contributing to protein anchorage. Interestingly, as mentioned above, the most conserved gene between TnGBS1 (*gbs0393*) and TnGBS2 (*gbs1143*) codes for an aggregation protein similar to the one found in *Enterococcus faecalis* pCF10 plasmid and in the sex factor of *Lactococcus lactis*. This protein was shown to promote efficient transfer of these conjugative elements. Furthermore this protein is also linked to streptococcal antigen I/II polypeptide family adhesins. A member of this protein family have been shown to be involved in tooth attachment, biofilm formation, tissue invasion and immune modulation in *Streptococcus mutans*.<sup>18</sup> This observation suggests that this protein may also confer a selective advantage in some conditions in addition to their role in the conjugative transfer of the TnGBSs. Finally, the ICE identified in *L. salivarius* carries a putative pilus operon, which may contribute to the adhesion of this bacterium or to the formation of biofilm as shown in other Gram-positive species.<sup>19</sup> Nevertheless, these ICEs are probably also selfish elements, imposing a limited fitness cost to the bacterial host as a result of their specificity of insertion in intergenic regions (see below).<sup>5</sup>

### Organization of ICE6013

ICE6013 is 13.3 kb long and related to the ICEBs1 family.<sup>20</sup> Its G+C content of 30.1% is slightly below the average G+C content of *S. aureus* genomes of 32.7%. A functional map is presented in Figure 1, highlighting the similarity with ICEBs1. However, the integrase encoding gene is replaced by a transposase. The relaxase gene is similar to the *nicK* gene encoded by ICEBs1.<sup>20</sup> The identification of such protein is indicative that ICE6013 may be conjugative. ICE6013 was identified in 6 out of the 15 strains of *S. aureus* sequenced at that time, where it was inserted at five different chromosomal *loci*. It was also shown to be widespread among *S. aureus* strains. In addition, these ICEs show a high level of polymorphism and signatures of recombination.<sup>6</sup> These observations strongly suggest that they are active and likely mobile. In the *S. aureus* ST239 strain where it was initially described, ICE6013 carries a Tn552 transposon encoding an ampicillin resistance gene. This gene was used as a marker for conjugation assays by using different Amp<sup>S</sup> EryR strains of *S. aureus* as recipient cells.<sup>6</sup> No transfer was detected in filter assays. However, in this ICE, Tn552 is inserted into the putative transposase gene, explaining these negative results. Therefore, whether this ICE is fully functional remains to be demonstrated.

ICE6013 was identified by PCR or by DNA gel blot in 19 out of 44 (43%) of clonally diverse strains of *S. aureus*. In most cases, these elements do not carry Tn552 or genes providing an obvious selective advantage. It seems therefore that this ICE spread among *S. aureus* genomes even in the absence of any cargo gene.<sup>6</sup> This ICE was also characterized independently within the SCCmec cassette of strain JCSC6826.<sup>21</sup> A more distantly related ICE was identified in the ED98 strain of poultry origin (GenBank: CP001781). This ICE is 75% identical to ICE6013 over its complete length, showing that these ICEs are probably ancient. In addition, similarity search at the NCBI nr nucleotide sequence database revealed a *S. aureus* plasmid (SAP077B, GenBank: GQ900429) almost identical to ICE6013. This plasmid corresponds to the circular intermediate of an ICE with a three nucleotide intervening sequence located between the two IRs: acaccacc\_aat\_ggcagtggtg. This organization is identical to the one described for the junction of the circular form characterized for ICE6013 expressing a full-length transposase. These three bases correspond to the three bases duplicated at the insertion site. It has been shown that ICEBs1 is capable of autonomous plasmid-like replication of its circular form after excision. The NicK protein, conserved in ICE6013, has been shown to be responsible for initiating the autonomous plasmid-like replication of ICEBs1.<sup>22</sup> It is therefore likely that ICE6013 may also be replicative under certain conditions as suggested for TnGBS elements.

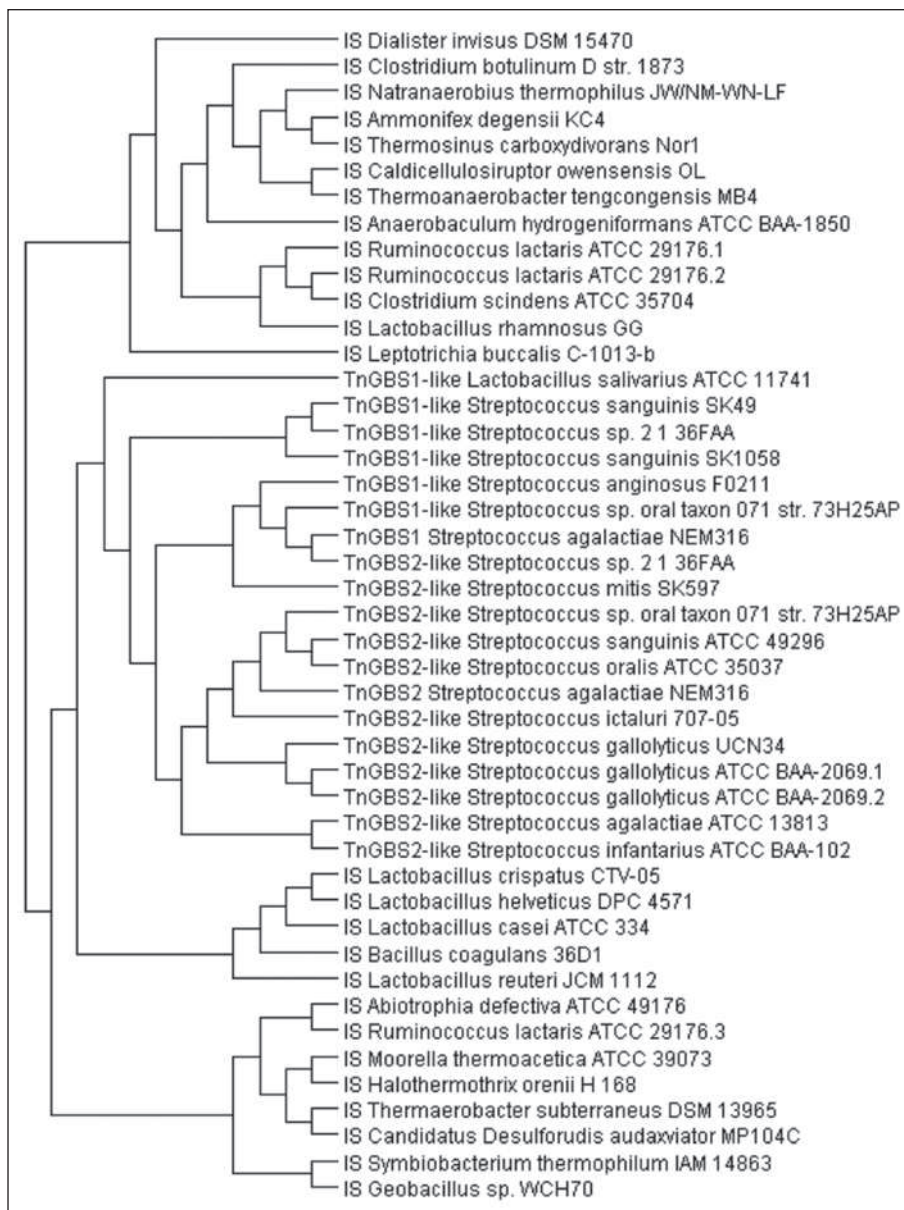


Figure 2. Unrooted phylogenetic tree of transposases from TnGBS-related elements and ISs. Species and strain names are indicated. Transposase sequences were retrieved at the NCBI databank after psi-blast search using TnGBS2 transposase as query. Protein sequences were aligned with MUSCLE using default parameters. The tree was built using the Jalview software with the average distance method and the BLOSUM62 matrix. Only a sample of divergent IS transposases representative of the major lineages were kept in the final alignment.

## The Transposition Module

Both Tn*GBS1* and Tn*GBS2* show a typical organization of transposable elements with two imperfect IRs at each extremity flanked by nine or ten base direct repeats corresponding to the duplication of the insertion site. Both elements carry at the right extremity a gene encoding a DDE transposase corresponding to a new sub-class. The transposases of both ICEs did not show significant similarity to any characterized transposase. However, motif searches at Pfam indicate that they belong to the uncharacterized protein family UPF0236 that is part of the Ribonuclease H superfamily which includes diverse families of DDE transposases. Blast searches revealed an ever-growing number of similar transposases mostly in Firmicutes. Some of these sequences are present in multiple copies in the host genome (up to 21 in *Caldicellulosiruptor owensensis* strain OL) and the analysis of the surrounding sequences revealed that they are also frequently flanked by inverted repeats and direct repeats of 7 to 10 bases.<sup>5</sup> These transposase genes together with the IR sequences thus represent a new family of Insertion Sequences (ISs). Phylogenetic analysis of transposases encoded by either ISs or conjugative transposons shows that Tn*GBS*-related transposases from streptococci cluster in a specific branch (R. Guérillot and V. Dacunha, unpublished). Despite the diversity of ISs in different species, each IS appears to be associated to a specific host (R. Guérillot and V. Da Cunha, unpublished).

Tn*GBS*-related transposases show a unique specificity among bacterial transposable elements. The ICEs (Table 1) and the ISs are almost systematically inserted in a defined orientation 15 or 16 bases upstream  $\sigma A$  dependent promoters.<sup>5</sup> This specificity, unique among prokaryotes, is reminiscent of that of eukaryotic retroelements, like for instance Ty3 from the yeast *Saccharomyces cerevisiae*. Ty3 inserts at the transcription start site of RNA polymerase III (PolIII) dependent promoters.<sup>23</sup> It has been shown that this targeting involves a direct interaction between the Ty3 integrase and the transcription factor TFIIB or TFIIC.<sup>24</sup> Similarly, the insertion specificity of the Tn*GBS* subfamily of transposases may rely on an interaction with the RNA polymerase. This specific interaction could have contributed to the observed restricted host range of these transposable elements, particularly if it involves variable parts of the RNA polymerase subunits. Although this insertion specificity is probably not essential for the coupling with conjugation, it has been shown that it correspond to “safe” localizations which does not affect significantly the expression of the downstream genes, minimizing the burden of the transposition on the bacterial recipient host.<sup>5</sup>

The transposase of ICE*6013* is related to the well characterized IS*30* family of transposases, but shows important differences. In particular it is split in two subunits and it misses the N-terminal helix turn helix motif conserved in IS*30* transposases.<sup>6</sup> In strains carrying an intact transposase gene, a circular form of the ICE*6013* was detected.<sup>6,21</sup> This observation is in agreement with a previous study demonstrating that a circular IS*30* element acts as an intermediate for simple insertion.<sup>25</sup> Interestingly, it has been demonstrated that IS*30* transposase from *Escherichia coli* can replace the recombinase function in an experimental lambda phage system.<sup>26</sup> This experiment reproduces the natural replacement observed between ICE*6013* and ICE*Bs1* types of ICE.

Unlike for the IS*30* transposase, the mechanism of transposition of Tn*GBS* transposases has not been characterized in depth. However, the transposition processes of Tn*GBS*-related transposases and of ICE*6013* are compatible with conjugation. In Tn*GBS1* and Tn*GBS2*, transposases catalyze the circularization of the transposon and the coupling sequence corresponding to the two IRs and the duplicated target could be amplified by PCR by using divergent primers located at both ends of the transposon.<sup>5</sup> A circular form of the ISs inserted into *Lactobacillus reuteri* genome was also detected by PCR and the coupling sequences corresponded to the four different insertion sites (V. Da Cunha, unpublished). Like for the integrase, this reaction produces a substrate for the conjugation machinery. However, Tn*GBS*s and IS-*L. reuteri* transposases do not restore the original insertion site of the element as opposed to integrases which catalyze a precise excision from the chromosome.. This suggests a replicative mechanism of transposition for Tn*GBS*-related transposases.<sup>5</sup>

## Conclusion

In 2008, the association of a conjugative module to a DDE transposase in Tn*GBS1* and Tn*GBS2* may have been anecdotal. However, a similar combination was reported one year later in *S. aureus* ICE*6013* and with the increasing number of sequenced genomes, these combinations happen to be much more frequent than anticipated at this time. Tn*GBS*s contribute to the conjugative transfer of chromosomal DNA by an HFr type mechanism and have a major role in the evolution of *S. agalactiae*.<sup>27</sup> It has been proposed that ICE*6013* may promote similar transfers in *S. aureus*.<sup>6</sup> Comparison of the organization of Tn*GBS1*, Tn*GBS2* and ICE*6013* revealed two conserved features important for their conjugative transfer: the formation of a circular intermediate in the transposition process and the capacity to transiently replicate this circular form. The latter property shows the close proximity of these ICEs and conjugative plasmids. Furthermore the general transfer process of these ICEs follows the same scheme described for other ICEs. The further characterization of these mobile elements will bring new insight into the diversity of ICEs. A key issue will be to understand the regulation and the coupling between the transposition process, the conjugation and the transient replication in the dissemination of these elements.

## Acknowledgments

This work was supported by the French National Research Agency (grant ANR-08-GENM-027-001) and by the Fondation pour la Recherche Médicale (FRM).

## References

1. Wozniak RA, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 2010; 8:552-63. PMID:20601965 doi:10.1038/nrmicro2382
2. Burrus V, Pavlovic G, Decaris B, Guedon G. Conjugative transposons: the tip of the iceberg. *Mol Microbiol* 2002; 46:601-10. PMID:12410819 doi:10.1046/j.1365-2958.2002.03191.x
3. Wang H, Roberts AP, Lyras D, Rood JJ, Wilks M, Mullany P. Characterization of the ends and target sites of the novel conjugative transposon Tn*5397* from *Clostridium difficile*: excision and circularization is mediated by the large resolvase, TndX. *J Bacteriol* 2000; 182:3775-83. PMID:10850994 doi:10.1128/JB.182.13.3775-3783.2000
4. Poyart-Salmeron C, Trieu-Cuot P, Carlier C, Courvalin P. Molecular characterization of two proteins involved in the excision of the conjugative transposon Tn*1545*: homologies with other site-specific recombinases. *EMBO J* 1989; 8:2425-33. PMID:2551683
5. Brochet M, Da Cunha V, Couve E, Rusniok C, Trieu-Cuot P, Glaser P. Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. *Mol Microbiol* 2009; 71:948-59. PMID:19183283 doi:10.1111/j.1365-2958.2008.06579.x
6. Smyth DS, Robinson DA. Integrative and sequence characteristics of a novel genetic element, ICE*6013*, in *Staphylococcus aureus*. *J Bacteriol* 2009; 191:5964-75. PMID:19648240 doi:10.1128/JB.00352-09
7. Tomita H, Ike Y. Genetic analysis of transfer-related regions of the vancomycin resistance Enterococcus conjugative plasmid pHTbeta: identification of oriT and a putative relaxase gene. *J Bacteriol* 2005; 187:7727-37. PMID:16267297 doi:10.1128/JB.187.22.7727-7737.2005
8. Smillie C, Garcillan-Barcia MP, Francia MV, Rocha EP, de la Cruz F. Mobility of plasmids. *Microbiol Mol Biol Rev* 2010; 74:434-52. PMID:20805406 doi:10.1128/MMBR.00020-10
9. Grohmann E, Muth G, Espinosa M. Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev* 2003; 67:277-301. PMID:12794193 doi:10.1128/MMBR.67.2.277-301.2003
10. Stentz R, Wegmann U, Parker M, et al. CsiA is a bacterial cell wall synthesis inhibitor contributing to DNA translocation through the cell envelope. *Mol Microbiol* 2009; 72:779-94. PMID:19400771 doi:10.1111/j.1365-2958.2009.06683.x
11. Wilcks A, Smidt L, Okstad OA, Kolsto AB, Mahillon J, Andrup L. Replication mechanism and sequence analysis of the replicon of pAW63, a conjugative plasmid from *Bacillus thuringiensis*. *J Bacteriol* 1999; 181:3193-200. PMID:10322022
12. Wickner S, Gottesman S, Skowrya D, Hoskins J, McKenney K, Maurizi MR. A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc Natl Acad Sci USA* 1994; 91:12218-22. PMID:7991609 doi:10.1073/pnas.91.25.12218
13. Smith MC, Thomas CD. An accessory protein is required for relaxosome formation by small staphylococcal plasmids. *J Bacteriol* 2004; 186:3363-73. PMID:15150221 doi:10.1128/JB.186.11.3363-3373.2004

14. Dunny GM. The peptide pheromone-inducible conjugation system of *Enterococcus faecalis* plasmid pCF10: cell-cell signalling, gene transfer, complexity and evolution. *Philos Trans R Soc Lond B Biol Sci* 2007; 362:1185-93. PMID:17360276 doi:10.1098/rstb.2007.2043
15. Brochet M, Couve E, Glaser P, Guedon G, Payot S. Integrative conjugative elements and related elements are major contributors to the genome diversity of *Streptococcus agalactiae*. *J Bacteriol* 2008; 190:6913-7. PMID:18708498 doi:10.1128/JB.00824-08
16. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 2010; 327:167-70. PMID:20056882 doi:10.1126/science.1179555
17. Rusniok C, Couve E, Da Cunha V, et al. Genome sequence of *Streptococcus galloyticus*: insights into its adaptation to the bovine rumen and its ability to cause endocarditis. *J Bacteriol* 2010; 192:2266-76. PMID:20139183 doi:10.1128/JB.01659-09
18. Brady LJ, Maddocks SE, Larson MR, et al. The changing faces of *Streptococcus* antigen I/II polypeptide family adhesins. *Mol Microbiol* 2010; 77:276-86. PMID:20497507 doi:10.1111/j.1365-2958.2010.07212.x
19. Konto-Ghiorgi Y, Mairey E, Mallet A, et al. Dual role for pilus in adherence to epithelial cells and biofilm formation in *Streptococcus agalactiae*. *PLoS Pathog* 2009; 5:e1000422. PMID:19424490 doi:10.1371/journal.ppat.1000422
20. Lee CA, Grossman AD. Identification of the origin of transfer (oriT) and DNA relaxase required for conjugation of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*. *J Bacteriol* 2007; 189:7254-61. PMID:17693500 doi:10.1128/JB.00932-07
21. Han X, Ito T, Takeuchi F, et al. Identification of a novel variant of staphylococcal cassette chromosome mec, type II.5, and its truncated form by insertion of putative conjugative transposon Tn6012. *Antimicrob Agents Chemother* 2009; 53:2616-9. PMID:19364875 doi:10.1128/AAC.00772-08
22. Lee CA, Babic A, Grossman AD. Autonomous plasmid-like replication of a conjugative transposon. *Mol Microbiol* 2010; 75:268-79. PMID:19943900 doi:10.1111/j.1365-2958.2009.06985.x
23. Chalker DL, Sandmeyer SB. Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev* 1992; 6:117-28. PMID:1309715 doi:10.1101/gad.6.1.117
24. Kirchner J, Connolly CM, Sandmeyer SB. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. *Science* 1995; 267:1488-91. PMID:7878467 doi:10.1126/science.7878467
25. Kiss J, Olasz F. Formation and transposition of the covalently closed IS30 circle: the relation between tandem dimers and monomeric circles. *Mol Microbiol* 1999; 34:37-52. PMID:10540284 doi:10.1046/j.1365-2958.1999.01567.x
26. Kiss J, Szabo M, Olasz F. Site-specific recombination by the DDE family member mobile element IS30 transposase. *Proc Natl Acad Sci USA* 2003; 100:15000-5. PMID:14665688 doi:10.1073/pnas.2436518100
27. Brochet M, Rusniok C, Couve E, et al. Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proc Natl Acad Sci USA* 2008; 105:15961-6. PMID:18832470 doi:10.1073/pnas.0803654105

# **Modular evolution and functional diversity of Integrative and Conjugative Elements of the Tn*GBS* family**

Running title: The Tn*GBS* ICEs family

<sup>1,2,3\*</sup>Romain Guérillot, <sup>1,2,\*</sup>Violette Da Cunha, <sup>1,2</sup>Elisabeth Couvé<sup>1</sup>, <sup>3</sup>Christiane Bouchier, and <sup>1,2,#</sup>Philippe Glaser

<sup>1</sup>Institut Pasteur, Laboratoire Evolution et Génomique Bactériennes, 28 rue du Dr Roux 75724 Paris; <sup>2</sup>CNRS *UMR* 3525 28 rue du Dr Roux 75724 Paris; <sup>3</sup>Université Pierre et Marie Curie, 2 Place Jussieu, 75005 Paris. <sup>4</sup>Institut Pasteur, and Genopole, Plateforme Génomique, 28 rue du Dr Roux 75724 Paris;

#To whom correspondence should be addressed

Philippe Glaser

Laboratoire Evolution et Génomique Bactériennes, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15 France. Tel: 33 1 45 68 89 96, Fax: 33 1 45 68 87 86

email: pglaser@pasteur.fr

\* These authors contributed equally to the work.



## Abstract

Integrative and conjugative elements (ICEs) have a major impact on gene flow and genome dynamics in bacteria. Unlike most ICEs characterized to date that use a phage like integrase for their mobility, TnGBSs use a DDE transposase. TnGBS1.1 and TnGBS2.2 in *Streptococcus agalactiae* strain NEM316 were the two first identified ICEs of this family. By systematic genome analysis we identified 56 different TnGBS-related ICEs, all but one being inserted in streptococcal genomes. These ICEs contain two types of conjugation modules, typified either by TnGBS1 or TnGBS2. Each of these conjugation modules, comprising a type four secretion system, a VirD4 coupling protein, a relaxase and its cognate *oriT* site, is shared with distinct families of Firmicute's conjugative plasmids. The TnGBSs also share replication modules with different plasmids and we show that both TnGBS1.1 and TnGBS2.1 combine replication and transposition to promote their spreading and stabilization. Phylogenetic analysis allows to reconstruct the successive recruitment by two conjugative elements of different origins of a transposition module probably derived from insertion sequences (IS). Despite an evolutionary successful horizontal spreading of TnGBS-related ICEs in the *Streptococcus* genus, these ICEs have a restricted host range. We show that for both TnGBS1 and TnGBS2, this host restriction is not due to a transfer incompatibility linked to the conjugation module, but most probably to their capacity to replicate in different hosts and/or to their transposition specificity upstream  $\sigma$ A dependent promoters, which is unique among bacterial transposable elements.

## **Introduction:**

Mobile genetic elements (MGEs) play a profound role in the evolution of both eukaryotes and prokaryotes. In bacteria, they are major contributors to gene flow within and among species and they may represent a significant proportion of their genomes. For example *Escherichia coli* O157H7 strain EC 4115 carries 18 integrated prophages totaling 786 kb covering 15% of its genome [1]. In the obligate intracellular bacterium *Orientia tsutsugamushi*, 30% of the genome is represented by the amplification of a 33kb Integrating and conjugative element (ICE) [2]. These elements may carry genes, which provide a selective advantage for their hosts such as specific metabolic functions, drug or heavy metal resistance as well as virulence factors. Indeed, MGEs were initially described as responsible for spreading antibiotic resistance [3]. However, systematic sequencing of bacterial genomes revealed a much broader diversity of MGEs, some without obvious selective function for the host.

Among the MGEs identified in bacterial genomes there is a growing interest for ICE [4,5]. These elements combine two properties, being conjugative like conjugative plasmids and having the capacity to integrate into the bacterial host genome like temperate phages. Integration into a host genome ensures the vertical transmission of these MGEs, whereas the capacity to be transferred by conjugation allows their horizontal spreading. In addition, ICEs have been shown to promote the conjugative transfer of chromosomal DNA among bacterial cells by the so-called Hfr mechanism [6]. Such chromosomal DNA transfers have been shown to contribute to chromosomal recombination among natural isolates of *Streptococcus agalactiae* [7], *Enterococcus faecalis* [8] and *Escherichia coli* [9]. A recent systematic search for conjugation machineries related to type four secretion systems (T4SS) in complete genome sequences, showed that ICE are the most abundant conjugative elements carrying T4SS among prokaryotes [5].

The conjugative transfer of an ICE requires its excision from the chromosome as a circular molecule. This reaction is catalyzed in most ICEs by a phage like integrase in some cases associated with an excisionase. Then, as for conjugative plasmids the DNA transfer is initiated by a relaxase, which performs a single strand nick of the circularized ICE at the *oriT* site. The single stranded DNA covalently bound to the relaxase is then recruited to a pore forming T4SS by the T4 coupling protein (T4CP) VirD4 where it is actively pumped into the recipient cell [10,11,12]. In the recipient cell the relaxase

catalyzes the circularization of the ICE, and the second strand is resynthesized leading to a double strand circular molecule, which is inserted into the chromosome of the recipient cell.

Most ICEs characterized, exemplified by Tn916, rely on an phage-type integrase for their mobility [13]. However, we recently described in *S. agalactiae* strain NEM316 two ICEs, TnGBS1.1 and TnGBS2.1 (initially named TnGBS1 and TnGBS2), that rely on a DDE transposase for their mobility [14]. TnGBS1.1 is 47 kb long, carries 49 protein encoding genes and is present in three copies, whereas TnGBS2.1 is 33.5 kb long and encodes 36 genes. These two ICEs have a G+C content of 37.6 and 38% respectively that is rather uniform along the sequence and slightly higher than the average value of the genome (35.6%). Like for Tn916, the conjugation process starts by the formation of a circular form of TnGBS2.1, which is the substrate of a relaxase that initiates the DNA transfer through a T4SS. In TnGBS2.1, the formation of this circular form is catalyzed by the transposase. However, unlike Tn916, the excision of TnGBS2.1 does not lead to the reconstitution of the insertion site in the donor strain. It was hypothesized that this first step of transposition is replicative. After conjugation the transposase catalyzes the integration of TnGBS2.1 into the chromosome of the recipient strain. Among bacterial transposable elements, these ICEs show a unique insertion specificity, 15 or 16 bases upstream the -35 region of  $\sigma$ A dependent promoters. This property is an intrinsic property of the transposition module constituted of the transposase and the two imperfect inverted repeats (IR) located at the ends of the transposable element [14].

At the time we have characterized these two ICE no similar element could be identified in publicly available bacterial genomes, and none was present in the seven other genomes of *S. agalactiae* available. Nevertheless, analysis by comparative genomic hybridization (CGH) of diverse *S. agalactiae* strains both of human and animal origin showed that these elements are frequent among strains of animal origin [14]. Furthermore, genes encoding proteins similar to these two transposases were identified in different Gram positive genomes, frequently in multiple copies. These genes were shown to correspond to a new family of widespread Insertion Sequences (ISs). These ISs show the same insertion specificity upstream  $\sigma$ A dependent promoters [14]. TnGBS1.1 and TnGBS2.1 share a similar transposition module, but encode different proteins involved in conjugation and relaxase modules and although TnGBS2.1 was shown to be efficiently transferred by conjugation among *S. agalactiae* strains, whether TnGBS1.1 is

mobile was not known. By systematic search in complete genome sequences, we have identified a large number of ICEs similar to either *TnGBS1.1* or *TnGBS2.1*. The *TnGBS* family represents therefore a new and broad family of ICEs that is almost exclusively found among the streptococcal genus. Using comparative analysis and phylogeny we have defined the functional modules involved in the integration, the conjugative transfer and the replication of these ICEs. *TnGBS* conjugative modules have two different origins but evolutionary converge by the recruitment of an unusual transposition module of IS origin for vertical maintenance. In addition, we show experimentally that *TnGBS1.1* and *TnGBS2.1* use a similar strategy involving a transient replication step in the recipient cell for enhancing their spreading prior to being stably inserted into the host genome. Experimental data, phylogeny and comparative analysis with different conjugative elements led us to construct a scenario for the generation of *TnGBS* ICEs in agreement with a general scheme for the evolution of conjugative elements unifying ICE and conjugative plasmid.

## RESULTS

### **TnGBS1.1 is conjugative like TnGBS2.1 and both have a narrow host range**

TnGBS2.1 can be efficiently transferred by conjugation from strain NEM316 to BM110 and subsequently from strain BM110 to other *S. agalactiae* strains [14]. In order to test whether TnGBS1.1 is also autonomously conjugative, we performed similar experiments by using strain GMP205 carrying an Erm cassette inserted in an intergenic region of TnGBS1.1. The marked TnGBS1.1 was transferred efficiently to strain BM110 (TetR) and subsequently retransferred to strain A909RF with a frequency of  $10^{-5}$  transconjugants (TC) per recipient strain, a value similar to the efficiency of transfer of TnGBS2.1 (Table 1).

The host range of ICEs and conjugative plasmids is highly variable. Tn916 is found in a broad range of Firmicutes [15] and can even be transferred from Firmicutes to Gram-negative bacteria [16]. In contrast the *E. faecalis* sex pheromone plasmids pAD1 and pIP1017 are restricted to this species [17]. We therefore wondered whether TnGBS1.1 and TnGBS2.1 could be transferred to other streptococci and to closely related Firmicutes. We performed conjugation assays using strain GMP205 and GMP201 as donor strains for TnGBS1.1 and TnGBS2.1, respectively and *Streptococcus gallolyticus*, *E. faecalis* and *Lactococcus lactis* strains as recipients (Table 1). TCs were obtained with *S. gallolyticus*, however with a  $10^4$  and  $10^3$  fold reduced frequency for TnGBS1.1 and TnGBS2.1, respectively. No TCs were obtained with *L. lactis* and *E. faecalis* in repeated experiments for both ICEs. Therefore under the conditions we have tested, both TnGBSs show a restricted host range for streptococci with a high preference for *S. agalactiae*.

### **TnGBS1.1 and TnGBS2.1 belong to a diverse family of streptococcal ICEs**

At the time of the first characterization of TnGBS1.1 and TnGBS2.1, no related ICEs were present in the sequenced genomes published [14]. Given the tremendous number of genome sequences and in particular of streptococcal species, available now we performed a new systematic search on all the whole genome sequences available at NCBI. We first searched by BLASTP and TBLASTN for genes encoding transposases similar to Gbs1118, the transposase of TnGBS2.1 and subsequently for proteins essential for transfer by conjugation, the relaxase, the VirD4 coupling protein the T4SS

component VirB4 located in the vicinity of the transposase gene. Finally we searched locally for the imperfect IR of the ICE by manual inspection of the surrounding sequences. At time of writing we identified 52 putative ICEs encoding a transposase of the TnGBS type. All are either related to TnGBS1.1 (21 ICE) or to TnGBS2.1 (31 ICEs) and several strains, like NEM316, contain two (six strains) or three (one strain) different ICEs of this family in their genome. All except one ICE were detected in streptococcal genomes (Figure 1). The only exception is a TnGBS1-like element that is inserted in the genome of *Lactobacillus salivarius* strain ATCC 11741. Genes similar to TnGBS1.1 genes were identified in two contigs of this unfinished genome sequence which extends did not allow the identification of the two IRs.

Among the 45 streptococcal species for which one or more genomes are available, 19 contain at least one TnGBS like ICE (Figure 1). The number of queried genomes per species ranges from 1 to 129 (for *S. pneumoniae*). TnGBS related ICEs are frequent in the genomes of *S. sanguinis*, *S. sobrinus*, *S. agalactiae*, in streptococci from the *bovis* group and from the *mitis* group. Interestingly, TnGBSs are frequent in *S. oralis* (2 out of 4 genomes) and in *S. mitis* (2 out of 11 genomes), but are present in only one of the 129 *S. pneumoniae* (strain GA17328) genomes sequenced. No ICE of this family was identified among the 15 *S. pyogenes* genomes published. For a majority of the streptococcal species only a single genome sequence is available, thus the fact that we identified a TnGBS like element in 19 species suggests that the spreading of this family of ICEs among streptococcal species is underestimated. To further explore the diversity of these ICEs we have completely sequenced four bovine *S. agalactiae* isolates using the Illumina technology and have determined the sequence of the four new TnGBS like ICEs encoded by these genomes. Taken together we have identified 36 complete and 20 partial ICEs of the TnGBS family (Table 2).

Some MGEs inserted into streptococcal genomes probably derived from TnGBS elements, as all sequenced *S. agalactiae* strains encode a conserved transposase similar to Gbs1118 even when no complete TnGBS element is present. Furthermore, in strain 2603V/R an IR-R repeat similar to that of TnGBS2 is present downstream the corresponding gene (SAG1237), however, no IR-L was detectable. This locus is located between a composite transposon carrying the genes *scpB* and *lmb* encoding the C5A peptidase and the laminin binding protein respectively [18] and a composite ICE inserted at the ribosomal protein *rplL*. This ICE is highly diverse in the different GBS

isolates [19]. Thus this transposase gene might represent a remnant of a *TnGBS* element that was inserted in a common ancestor of the modern GBS isolates. We also identified in *Streptococcus dysagalactiae* subsp. *equisimilis* a 6.3 kb transposon bordered by two IRs and expressing a transposase that is 91 % identical to the transposase expressed by *TnIctaluri-2-1* (Table 2). This transposon expresses proteins similar to proteins identified in other MGEs but is devoid of functions related to conjugation (data not shown).

*TnGBS1.1*, *TnGBS2.1* and ISs expressing a similar transposase are specifically inserted upstream  $\sigma$ A dependent promoters [14]. By analyzing their insertion sites, we showed that the newly identified ICEs show the same insertion specificity as all 51 ICEs whose insertion site was identified are inserted upstream  $\sigma$ A dependent promoters (Table 2). The targeted genes are in great majority conserved house keeping genes. In particular 20 ICEs are inserted upstream genes encoding ribosomal proteins. Nine genes were targeted by different ICEs independently, e.g. *rpmH* was targeted four times. Only the promoter region upstream *pyrG* and *rpmV* were targeted by both *TnGBS1* and *TnGBS2* related elements.

In order to determine the shared features amongst *TnGBS1* and *TnGBS2* related elements, we selected those ICEs that were complete or almost complete and that were representative of the diversity of all identified ICEs for further analyses. We obtained 33 ICEs for which we performed systematic BLASTP and/or BLASTX comparisons against the protein complements of *TnGBS1.1* or *TnGBS2.2*. Indeed, proteins encoded by these ICEs show higher similarity either with *TnGBS1.1* or with *TnGBS2.1* proteins (Figure 2 and 3). Therefore ICEs belonging to the *TnGBS* family can be classified into two subtypes exemplified by *TnGBS1.1* and *TnGBS2.1* and were thus named accordingly (see legend of Table 2). We identified 34 proteins shared by the 10 streptococcal ICEs related to *TnGBS1.1* and 14 proteins shared by the 23 *TnGBS2.1* related elements. These core sets of conserved genes encode proteins likely required for the mobility of these ICEs as several are similar to proteins involved in the conjugation of different plasmids from Firmicutes as depicted in Figure 2 and 3. This finding reflects the proximity of *TnGBS*s to two different families of Gram-positive conjugative elements (Figure 2 and 3). In particular, proteins essential for conjugation like the coupling protein VirD4 and components of the T4SS are among the most conserved proteins as well as the relaxase and its accessory protein. Interestingly, we also observed conserved proteins involved in

plasmid replication. Taking together this information, we grouped these genes into different functional modules for conjugation, replication and transposition (Figure 2 and 3). Genes that were only present in one ICE or in few more similar ICEs may encode accessory proteins of these modules or may represent cargo genes with functions not directly involved in mobility (Figure S1 and S2). In the following sections, we have performed a comparative analysis of the different functional modules of Tn*GBS* ICEs.

### **Tn*GBS*s arise by recruitments of a transposition module of IS origin by unrelated conjugative elements**

The common feature defining the Tn*GBS* family is a shared transposition module. We performed a phylogenetic analysis of the complete transposase sequences of 44 Tn*GBS*-like ICEs with the transposases of the related IS from *Lactobacillus helveticus*, *L. crispatus* and *Bacillus coagulans* as an out group (Figure 4). As deduced from this analysis depicted in Figure 4, the transposases of the streptococcal ICEs represent a monophyletic group within a widespread family of ISs (data not shown) indicating a common IS origin for this transposition module. All the transposases of the Tn*GBS*1 group, except the ill characterized Tn*GBS*-*Lsaliv*1 represent by themselves a monophyletic group branching within the transposase of Tn*GBS*2 like ICEs. Therefore, except the ICE identified in *L. salivarius*, Tn*GBS* ICEs probably have a common origin of a Tn*GBS*2.1 type. The streptococcal Tn*GBS*1s might have emerged among Tn*GBS*2 like ICEs by exchanging the modules involved in its conjugation. This transposition module being therefore recruited by two unrelated conjugative elements.

The second element of the transposition module is represented by the two IRs. The comparison of the IRs of 41 ICEs defines two well-resolved groups with two different consensus sequences (Figure 5). The IR-R located downstream the transposase gene is almost completely conserved within each group. Strikingly these two groups correspond to ICEs either related to Tn*GBS*1.1 or to Tn*GBS*2.1 (Figure 5). Seven streptococcal isolates carry both a Tn*GBS*1.1 and a Tn*GBS*2.1 like element. The difference between the IRs of these two elements may prevent the crossrecognition of the IRs of an ICE by the transposase of the other ICE. Indeed, the GMP202 strain in which the transposase gene has been inactivated does not produce a circular form of



Tn*GBS*2.1 although it expresses three copies of the Tn*GBS*1.1 transposase, which is in support to this hypothesis [14].

The Tn*GBS*2.1 transposition process involves two steps: the formation of a circular form in the donor strain and the integration into the genome of the recipient strain [14]. We have shown that these two steps in Tn*GBS*2.1 are catalyzed by the transposase encoded by *gbs1118*. In order to test whether the putative Tn*GBS*1.1 transposase encoded by *gbs0410* is also sufficient for transposition, we have constructed a non replicative plasmid mimicking the circular form of Tn*GBS*1.1 (pSU-*erm*::miniTn*GBS*1, Figure S3) and tested its transposition. This pSU18 derivative carries an *erm* gene as selective marker and the transposase gene *gbs0410* together with the two IRs. After electrotransformation into the *S. agalactiae* strains BM110 and NEM316, transformants were obtained. These transformants harbor an inserted copy of pSU-*erm*::miniTn*GBS*1 as shown by direct sequencing of the plasmid-chromosome junctions (see below). *Gbs0410* is therefore sufficient for the insertion of the transposon into the recipient chromosome. We then tested by PCR for the presence of a circular form of the ICE both in strains carrying Tn*GBS*1.1 (NEM316 and GMP105) and in pSU-*erm*::miniTn*GBS*1 transformants (Figure S3). We were able to amplify the junction sequence of Tn*GBS*1.1 and pSU-*erm*::miniTn*GBS*1 by using divergent primers located at both ends of the ICE (O-11, O-12) (data not shown). This sequence corresponds to the two IRs separated by a 7 to 10 nt sequence which corresponds to the sequence duplicated at the insertion site. Therefore *Gbs0410* is also sufficient for the formation of a circular form of Tn*GBS*1.1. The formation of a circular form of Tn*GBS*2.1 does not correspond to an excision restoring the empty insertion site. PCR amplification using the oligonucleotides pairs O-7 / O-8 and O-9 / O-10 failed to amplify the insertion site in strains NEM316 and GMP205, respectively. This observation shows that the formation of a circular intermediate of Tn*GBS*1.1 does not rely on the excision of the ICE and supports the hypothesis of a replicative mechanism for the transposition of both subtypes of Tn*GBS*.

Tn*GBS*2.1 inserts specifically upstream the -35 sequence of  $\sigma$ A dependent promoters [14]. Although it targets different promoters the *nrdF* gene (*gbs0453*) encoding the small subunit of the nucleotide reductase is a preferential insertion site. The three Tn*GBS*1.1 copies in strain NEM316 are also inserted 15 to 17 bases upstream the  $\sigma$ A promoters of *gbs0411*, *gbs0741* and *gbs0968*, respectively [14]. In order to

determine the specificity of the Tn*GBS1.1* transposase we first attempted to determine the insertion site of Tn*GBS1.1* into BM110 TCs. However, in all the TCs analyzed directly after conjugation, the DNA sequence obtained corresponded to the circular form of the transposon. Moreover, the sequence located between the two IR corresponded to the insertion site in the GMP205 donor strain (see below). Nevertheless, after up to 7 serial passages on TH agar plates, we identified 16 clones with a copy of Tn*GBS1.1* inserted in the genome. The 16 Tn*GBS1.1* were inserted upstream  $\sigma$ A dependent promoters, and only two promoters, upstream *rpsJ* and *rpsU* were targeted by two independent insertions (Figure S4). In order to characterize a larger number of insertion sites we sequenced the insertion sites of the pSU-*erm::miniTnGBS1* in strains NEM316 and BM110. All insertions tested were located upstream promoters in the same relative orientation as Tn*GBS2.1*: the IR-R 15-17 bp upstream the -35 region of a  $\sigma$ A dependent promoter (Figure S4). The analysis of 100 insertion sites identified 56 different sites and thus only a low level of preference for specific promoter regions (Figure S4). Strikingly, comparison of the promoters targeted by Tn*GBS1.1* and Tn*GBS2.1* identified a single shared position that is upstream *gbs0411*. In particular we did not identify any insertion of pSU-*erm::miniTnGBS1* upstream *nrdF*, the preferred insertion site of Tn*GBS2.1*. Therefore the process of promoter targeting by the two ICEs is probably conserved but they show subtle differences leading to differences in insertion specificity.

### **Replication of Tn*GBS1.1* and Tn*GBS2.1* in the recipient strain promotes its transfer**

The identification of Tn*GBS1.1* TCs carrying a circular form of the ICE and the absence of an integrated copy provide a first hint that this ICE may have the capacity to replicate before its insertion in the recipient cell. Analysis of the functions encoded by Tn*GBS1* and Tn*GBS2* like elements revealed the presence of conserved plasmid related replication modules. All Tn*GBS2* like ICEs encode a RepA\_N like protein. Similar proteins are also encoded by the plasmids pCF10, pTEF2, pSK41 and pBD-II and by the ICEs Tn5253, ICEde3396 and the *S. suis* 89K pathogenicity island (Figure 4). RepA\_N initiates the replication of a broad range of plasmids from different Firmicutes [20]. These plasmids show a restricted host specificity that has been attributed to the specificity of the RepA\_N protein. The origin of replication targeted by RepA\_N in these plasmids lies within the coding sequence of *repA*. These *ori* sequences are generally not conserved in

sequence but share similar features including direct and inverted repeats [20]. By analyzing the Tn*GBS2.1* *repA* nucleotide sequence we identified a sequence similar to the *oriR* region of the *Staphylococcus aureus* plasmid pSK41 [21] (Figure 6A). This sequence represents a putative origin of replication recognized by Tn*GBS2.1* RepA protein. Other proteins (RepB and RepC) shown to be involved in plasmid stability [22] are missing in Tn*GBS2.1* and in related elements.

Tn*GBS1.1* encodes also a RepA\_N like protein (Gbs0408). However, this protein is only 127 aa long. Furthermore orthologs of this gene are missing in 4 of the 11 Tn*GBS1*-like elements analyzed and it is thus perhaps not the main replication protein of Tn*GBS1.1* (Figure 3). However, we identified a second replication module, encoding a RepE homolog that is present in all Tn*GBS1*-like ICEs. The RepE protein is essential for  $\Theta$ -replication of the broad host range plasmid pAM $\beta$ 1 and binds to the origin of replication located downstream the *repE* gene [23,24]. By analyzing the DNA region surrounding the *repE* gene (*gbs0367*) in Tn*GBS1* like elements, we identified a conserved sequence similar to the origin of replication characterized in plasmids pAM $\beta$ 1, pSM19035, pIP501 and pXO1 [25] (Figure 6B). In contrast, like for the Tn*GBS2* replication module, we did not identify the *seg* locus involved in the stability of these plasmids [26]. Nevertheless, additional proteins conserved among all Tn*GBS1* like ICEs might contribute to plasmid replication like a type 1 topoisomerase that is also conserved in related plasmids (Figure 3). Indeed an homolog of type 1 topoisomerase was shown to influence pAM $\beta$ 1 replication [24]. In addition, we identified in all Tn*GBS1*-like elements and in eight Tn*GBS2*-like elements a gene encoding a protein similar to the ParA protein from the *Lactococcus lactis* plasmid pCI2000, which is involved in plasmid stability [27]. Compared to Tn*GBS2*-like ICEs, Tn*GBS1*-like ICEs seem therefore to encode a more complete set of genes involved in plasmid replication.

In order to further demonstrate that replication of Tn*GBS1.1* and Tn*GBS2.1* takes place in the recipient strain we analyzed the mark left in the circular form between the two IRs, which corresponds to the duplicated sequence at the insertion site. Our reasoning was that before their insertion in the recipient strain the coupling sequence of the transferred Tn*GBS1.1* and Tn*GBS2.1* contain the nucleotides duplicated at the insertion site in the donor strain, whereas the circular form generated from an inserted copy in the recipient strain carries the sequence corresponding to the new insertion site. We thus PCR amplified, cloned and sequenced the coupling sequences of TCs from

Tn*GBS*1.1 and Tn*GBS*2.1 (Table 3). In all Tn*GBS*1.1 sequence analyzed, the coupling sequences corresponded to transferred circular ICE, confirming the replication of Tn*GBS*1.1 in the recipient strain. In the case of Tn*GBS*2.1, 97.5% of the circular form corresponded to transferred circular Tn*GBS*2.1 (Table 3). We also observed few coupling sequences corresponding to inserted forms in the recipient. Five out of seven were inserted in *nrdF*, the preferential site of Tn*GBS*2.1. As the frequency of circularization of an integrated Tn*GBS*2.1 is low (less than  $10^{-3}$ ), the percentage of these coupling sequences reflects a significant proportion of cells with an integrated copy. The identification of circular forms corresponding to inserted copies indicates that the delay in the insertion of the ICE in the recipient chromosome is shorter for Tn*GBS*2.1 than for Tn*GBS*1.1.

In order to quantify the replication of the two ICEs we determined the ratio of circular forms compared to the total amount of Tn*GBS*1.1 and Tn*GBS*2.2 in the donor strain and in TCs by Q-PCR (Table 4). We compared the number of copies of the coupling sequence of the ICE and of the transposase gene and normalized our results with the chromosomal gene *polA*. Similarly to what we previously reported, the ratio between circular and inserted Tn*GBS*2.1 forms was  $8 \cdot 10^{-4}$ . In the case of Tn*GBS*1.1 (strain GMP205) the ratio was about 100 times higher than for Tn*GBS*2.1. For both ICEs, we observed a strong increase in the number of copies of the circular form in TCs. In the case of Tn*GBS*1.1 we observed about one circular form per chromosome, a value compatible with a replication of Tn*GBS*1.1 and a control of the copy number. In the case of Tn*GBS*2.1 we observed only 0.26 copies per chromosome. This value reflects a low efficiency of replication and the insertion of the ICE into the chromosome of a significant proportion of the cells. To test whether the increased number of circular forms might lead to an increased efficiency of transfer, we performed conjugation assays by using these Tn*GBS*1.1 and Tn*GBS*2.1 TCs as donor strains (Table 4). Indeed, in both cases we observed a strong increase of the conjugation efficiency. These results suggest that the frequency of transfer is limited by the number of circular forms in the donor strains and that replication after conjugation is a means to increase transiently the efficiency of transfer.

### **Tn*GBS*1 and Tn*GBS*2 like ICEs express two distinct conjugation modules**

Similarity searches with conjugative elements show that both Tn*GBS*1 and Tn*GBS*2-like ICE encode homologs of VirD4 and VirB4 and of relaxases from different plasmids and ICEs (Figure 2 and 3). In order to understand how these modules were acquired, we performed a comparative phylogenetic analysis of VirB4, VirD4 and the relaxase proteins. We compared Tn*GBS*1 type proteins with homologous proteins from conjugative plasmids. We observed a congruence of the three trees in agreement with a single origin for the conjugation / mobilization module (Figure S5). Interestingly in the case of *L. salivarius*, although the genes are on two different contigs, the topology of the tree is similar, confirming the hypothesis that these two contigs correspond to the two parts of a Tn*GBS* related ICE that have acquired a conjugation / mobilization module similar to Tn*GBS*1 (Figure S5). Similarly, we observed a congruence in the phylogeny of of these three genes for Tn*GBS*2 related ICEs, conjugative plasmids of Firmicutes and widespread ICEs from Firmicutes depending on a phage like integrase (Figure S5). The congruence and the similarity of the trees, indicate a single origin of the conjugation/relaxase module of Tn*GBS*2. However, for Tn*GBS*2 the phylogeny of the transposase and that of proteins of the conjugation/relaxase module show several cases of incongruence indicating exchanges of the conjugative module among Tn*GBS*2 related ICEs.

When trying to identify the relaxase of Tn*GBS*1 we first could not identify any protein encoding a known relaxase domain. However, since we have demonstrated the autonomous mobility of Tn*GBS*1, we further searched for a possible relaxase by considering conserved proteins of unknown function among all Tn*GBS*1 related ICE. Indeed, PSI-BLAST search based on Gbs0384 gave numerous hits in conjugative plasmid relaxases. Alignment of Gbs0384 and its distantly related homologs allowed us to identify the characteristic conserved residues of relaxases in the N-terminal part of the protein: two potential catalytic Y and a 3H motif [28]. Furthermore, distant homologs of the N-terminal part of Gbs0384 have been proposed to belong to a new family of relaxases responsible for the mobility of the vancomycin resistance conjugative plasmid in *Enterococcus* pHT $\beta$  and pMG2200 [29].

Next we tried to identify the origin of transfer. The functional *oriT* sequence of pMG2200 has been identified [30]. We therefore searched for sequences with a similar organization in Tn*GBS*1-like ICEs. Indeed, we identified a conserved sequence sharing similar features in all Tn*GBS*1-like representatives: two inverted repeats followed by a

20 bp conserved sequence (Figure 7). This sequence is located upstream *gbs0380* that probably represents the first gene of the mobilization module of Tn*GBS1.1*. Genes similar to the four genes of this locus were identified in all Tn*GBS1* related ICEs. In order to test the functionality of this putative  $oriT_{TnGBS1}$  site, we cloned a one kb fragment encompassing this region into the non mobilizable vector pTCVerm [31] and tested for the conjugative transfer of the resulting plasmid from NEM316 to BM110. Indeed, this plasmid was efficiently transferred to strain BM110 demonstrating that this region encompassed the origin of transfer (Table 5).

In Tn*GBS2.1* *gbs1121* shows a strong conservation among all Tn*GBS2*-like ICEs and is homologous to the relaxase of the pheromone responsive plasmid pCF10 from *E. faecalis* and pRS01 from *L. lactis* [32]. They all belong to the MOB<sub>P7</sub> clade [33]. *gbs1121* was shown to be essential for the transfer of Tn*GBS2.1* [14]. The activities of pCF10 and pRS01 relaxases on their cognate *oriT* region have been biochemically and genetically confirmed [32]. In both plasmids the recruitment of the relaxase on the double stranded *oriT* requires a small accessory protein (PcfF in pCF10 and LtrF in pRS01). Interestingly, this protein is present and well conserved (31% and 33% identity respectively) among all Tn*GBS2* related ICEs (*gbs1122* in Tn*GBS2.1*). In Tn*GBS2.1* an *oriT* region was previously identified between the IR-R and the transposase gene [14]. However, this *oriT* is unrelated to the one described in pCF10, pRS101 and pC221. Comparison of the DNA sequence upstream *gbs1123* revealed a c.a. 100 bp sequence conserved among Tn*GBS2*-like ICEs. It is also reminiscent of the *oriT* of pCF10 and pRS01 and of other MOB<sub>P7</sub> clade members (Figure 8). In particular, this secondary *oriT* present a conserved organization with an inverted repeat upstream a short conserved sequences (Figure 8). These conserved sequences were biochemically demonstrated to include the nick site of the relaxase of pCF10, pRS01 and pC221 [32] and the corresponding position represents therefore the likely nick site for Tn*GBS2* (Figure 7). The prediction of this putative second origin of transfer, prompted us to test its functionality. The corresponding DNA region from Tn*GBS2.1* was cloned into pTCVerm and the resulting plasmid was essayed for conjugative transfer from strain NEM316 to strain BM110 (Table 5). This experiment showed that this region carries a genuine origin of transfer that promotes the efficient transfer of pTCVerm among *S. agalactiae* strains.

pTCVerm can replicate in diverse Firmicutes, in order to test whether the restricted host range of TnGBS1.1 and TnGBS2.1 was due to restriction in the mobilization or conjugation steps, we tested for the transfer of the two pTCVerm derivatives carrying *oriT*<sub>TnGBS2</sub> and *oriT*<sub>TnGBS1</sub> respectively to *E. faecalis*. For both plasmids we observed a similar efficiency of conjugation from NEM316 to *E. faecalis* and from NEM316 to the *S. agalactiae* strain BM110 (Table 5). This result demonstrates that the narrow host specificity of TnGBS elements does not result from a specificity of the conjugation machinery.

## DISCUSSION

The analysis of the complete genome sequence of diverse streptococcal isolates belonging to 44 species revealed that despite the absence of obvious cargo genes, ICEs related to TnGBS1 and TnGBS2 represent a broad family of conjugative elements. Among the experimentally described ICEs [13], TnGBSs are as diverse as the Tn916 family in Firmicutes which express the tetracycline resistance gene TetM and other antibiotic resistance genes like *erm*(B) and *aph*-A3 [34] or the SXT/R391 family from Gram negative bacteria that express diverse antibiotic resistance genes and/or other specific functions [35]. The absence of typical cargo genes such as antibiotic or heavy metal resistance explains why despite the expansion of this family among streptococcal species they were only discovered through systematic genome sequencing [36]. These ICEs might be just parasitic elements. We have shown that, their unique specificity in intergenic regions upstream the -35 region of canonical  $\sigma$ A dependent promoters minimizes the fitness cost of their insertion [14]. Nevertheless, some functions encoded by these ICEs might affect the phenotype of their host providing a selective advantage or disadvantages under certain circumstances. In particular, the putative aggregation proteins (Gbs0393 and Gbs1143 expressed by TnGBS1 and 2 respectively) are related to streptococcal antigen I/II polypeptide family adhesins. In streptococci, members of this family have been shown to be involved in host colonization, biofilm formation and immune modulation [37,38]. Strikingly, TnGBS ICEs are underrepresented among isolates of human origin. Among the eight sequenced human *S. agalactiae* isolates, only NEM316 carries TnGBS elements and its human origin has been questioned [39]. The recently released sequence of two bovine isolates, ATCC 13813 (NZ\_AEQQ00000000) and FSL S3-026 [40] contains two and three TnGBS ICEs, respectively. TnGBSs are

frequent among the commensal species *S. oralis* and *S. mitis* whereas they are virtually absent from *S. pneumoniae* and completely absent from all *S. pyogenes* genomes available. It seems therefore that despite their mobility TnGBSs are counter-selected in human pathogenic streptococci. A possible reason might be that surface proteins expressed by these ICEs are recognized by the human immune system and contribute to the clearance of the pathogen.

Conjugative elements, plasmids and ICE have been shown to share numerous features and to exchange different modules [5,13]. TnGBS ICEs express two different conjugative modules related to conjugative plasmids. The global congruence of the phylogenetic trees for the VirD4, VirB4 and relaxase proteins indicates the *en bloc* exchange of these modules. *ser/tyr* recombinase dependent ICEs represent a broad family of mobile genetic element combining diverse integrases and conjugative modules [13]. Despite the diversity of transposases [41], this association observed in TnGBS ICEs is rare. TnGBS elements are only associated with two types of T4SS genes. The only other known case of such an association of a DDE transposase and a conjugative module has been recently reported for ICE6013 of *S. aureus*. ICE6013 expresses an IS30-like transposase and carries a conjugative module similar to that of ICEBs1 [42,43]. Interestingly, it has been demonstrated that IS30 transposases from *Escherichia coli* can replace the recombinase function in an experimental lambda phage system [44]. This experiment reproduces the natural replacement observed between ICE6013 and ICEBs1 types of ICE. A shared feature of the transposase of TnGBS elements and of the IS30 family is the formation of a circular intermediate that is required for the conjugation process [45]. However, different transposases like those from the broad IS3/IS911 family generate such circular intermediates [46], but have not been shown to be associated with a conjugative module to form an ICE. One could speculate that the conversion of a plasmid to an ICE could have been frequent, resulting from the Campbell type recombination of a plasmid at a conserved IS sequence to generate as a first step a composite conjugative transposon [47]. As such combinations are rare, it is likely that this association imposes additional constraints on both the transposition and the conjugative processes.

Comparative analysis of ICEs shows that the same mobilization/conjugation module may use different strategy for its stabilization: the replication as plasmid or the integration into the host chromosome [5,13]. TnGBS ICEs are stably maintained



integrated into the genome as the conjugative process does not involve the excision of the ICE which may lead to its loss as observed for ICE using a phage like integrase [48]. However, we have shown that all Tn*GBS* ICEs encode replication proteins and that both Tn*GBS*1.1 and Tn*GBS*2.1 are able to replicate in the recipient cell. Interestingly these replication modules are similar to the replication machinery of the related conjugative plasmids (Figure 2 and 3). We observed that accessory proteins involved in replication control and plasmid stability are missing. Therefore the capacity to replicate was not kept for the long-term stabilization of these ICEs and this replication is transient before integration. Nevertheless, this replication step is probably required for the integration to occur as a mutant of Tn*GBS*1.1 in *gbs0367* (*repE* like) failed to conjugate (data not shown). The balance between the replication and transposition process should be tightly regulated to limit the burden imposed by the ICE on its host. Although ICEs were generally thought not to replicate, replication has been reported for different ICEs [49]. It has been recently shown that ICE*B*s1 replicates in the donor strains. Replication was initiated by the ICE encoded relaxase NickK and it required the host DNA polymerase PolC. Unlike Tn*GBS*1.1, the replication of ICE*B*s1 is not required for its conjugative transfer but contributes to its stability [50] and also probably to the spreading of the ICE to many cells in chains after the initial conjugative transfer to one of the cells of the chain [51]. Replication of Tn*GBS* ICEs is different from that of ICE*B*s1 as it uses a typical replication machinery of Firmicute plasmids.

Our new results allow to propose the following model for the propagation and the stabilization of Tn*GBS* ICEs. As an integrated form, the ICE is inactive, thus leading to no burden for the host and it is stably transmitted to its descendants. After stochastic or regulated formation of the circular form, the ICE might express conjugative functions and might be efficiently transferred to a recipient cell, which would be highly proficient to replicate and retransfer the ICE to secondary recipient bacteria. After integration into the host genome, the ICE becomes silent again and is stably maintained.

Using conjugation assays under laboratory conditions, we have shown that Tn*GBS* could be efficiently transferred among *S. agalactiae* isolates, the efficiency of transfer dropped in other streptococcal species and no transfer was detected in closely related Firmicutes like *Enterococcus* or *Lactococcus*. This host specificity does not result from a limitation of the transfer as the conjugative module of both Tn*GBS*1 and Tn*GBS*2 promotes the transfer to *E. faecalis* of a non-mobilizable plasmid carrying their

putative *oriT* sites respectively. It has been shown that the host specificity of pAD1 and pIP1017 for *E. faecalis* results from the specificity of the RepA\_N protein [17]. The species specificity of replication may also contribute to the host specificity of Tn*GBS* ICEs. Supporting this hypothesis was the observation that in the few *S. gallolyticus* TCs obtained, Tn*GBS*1.1 was inserted into the chromosome. Tn*GBS*1.1 is probably unable to replicate in this streptococcal species suggesting also that the transposition process in the recipient strain required this initial replicative phase. However, this restriction may also result from the specificity of the transposition module. Indeed, after electroporation in *E. faecalis* pSU-erm::miniTn*GBS*2 has lost its insertion specificity upstream  $\sigma A$  dependent promoters and inserted with a low efficiency (data not shown). The insertion specificity and hence the mechanism of insertion may result from a tight co-evolution of the transposition module with the host transcription machinery leading to a drop of the transposition efficiency in non streptococcal hosts.

By similarity to what has been shown for eukaryotic retroelements such as Ty3 in yeast [52], we have proposed that the insertion specificity of Tn*GBS* ICEs rely on a specific interaction of the transposase and the  $\alpha$  or the  $\sigma$  subunit of the RNA polymerase during the transcription imitation step [14]. Our comparative analysis provides further clues on this puzzling specificity. All Tn*GBS* elements shared a similar specificity and among natural isolates they share some preferential sites upstream ribosomal protein-encoding gene. However, comparison of the insertion sites targeted by Tn*GBS*1.1 and Tn*GBS*2.1 revealed that the transposase together with the IR sequences targets different promoter regions. According to a model of interaction with RNA polymerase subunits, differences in the protein sequence of the transposase, might lead to different interactions with the RNA polymerase during the initiation of transcription leading to differences in preferred promoters. However, we observed striking differences between the IRs of the two families of transposases. In particular the TAGACA sequence in the highly conserved IR-L sequence of Tn*GBS*1 ICE is replaced by a TACCAC in Tn*GBS*2 related ICEs. The similarity of this sequence with the consensus of the -35 sequence might contribute to the insertion specificity and hence to the difference of specificity between the two ICEs.

We have shown that Tn*GBS*2.1 promotes in *S. agalactiae* the conjugative transfer of large chromosomal regions of up to more than 300 kb and contributes to shaping the genomic diversity of this species [7]. The discovery of Tn*GBS* ICEs in half of the

streptococcal species led us to speculate that they may similarly contribute to the evolution of these species by initiating Hfr transfer of chromosomal DNA. In particular the discovery of such ICEs in the *mitis* group and in one pneumococcal isolate poses the question of their possible role in chromosomal exchanges in this group of great clinical importance. This intriguing family of ICEs has been identified by analyzing complete genome sequences. It is likely that other families of ICEs remain to be discovered that might have similarly an important role in bacterial evolution and biology.

## **MATERIALS and METHODS**

**Bacterial strains and growth conditions.** Bacterial strains used in this work are listed in table S1. Conjugation assays were performed by using strain GMP201 and GMP205 as donor strains and strains BM110, BM110RF and A909RF as recipient strains. *Escherichia coli* strain DH5 $\alpha$  (Invitrogen) was used for subcloning. *S. agalactiae* was grown in Todd-Hewitt (TH) broth or agar (Difco Laboratories, Detroit, MI) and *E. coli* in Luria-Bertani medium. Antibiotics were used at the following concentrations: for *E. coli* – ampicillin, 100 mg l<sup>-1</sup>; erythromycin, 150 mg.l<sup>-1</sup>; for *S. agalactiae* – erythromycin, 10 mg.l<sup>-1</sup>; rifampicin, 100 mg.l<sup>-1</sup>; fusidic acid, 10 mg.l<sup>-1</sup>; tetracycline, 10 mg.l<sup>-1</sup>.

**General DNA techniques and sequencing.** Plasmid DNA was prepared by alkaline lysis using the Qiagen Miniprep plasmid kit and genomic DNA was extracted using the Qiagen DNeasy kit (Qiagen, Hilden, Germany). The oligonucleotide primers used for PCR reactions were designed with Primer3 software (<http://primer3.sourceforge.net>). To determine insertion sites of TnGBSs and of derived recombinant transposons, sequencing reactions were directly performed on genomic DNA. Complete genome sequences were determined by using the Illumina technology (GAIIIX). The 36 nucleotides reads were assembled by using the Velvet software [53]. Contigs corresponding to TnGBSs elements were identified by BLASTN and TBLASTN searches with the sequences of TnGBS1.1 and TnGBS2.1 and the sequences of the proteins encoded by these two ICEs as queries respectively. Sequences between contigs were determined by Sanger sequencing after PCR amplification of the corresponding regions. The four newly identified ICEs were annotated manually by similarity with TnGBS2.1 and submitted to EMBL (accession number pending).

**Plasmid and strain construction:** Oligonucleotides used in this work are listed in table S2. Mobilization plasmids carrying the putative *oriT* regions of Tn*GBS1.1* and Tn*GBS2.1* are derived from the non mobilizable broad spectrum vector pTC*Verm* [31]. The Tn*GBS1.1* and Tn*GBS2.1* *oriT* regions were amplified by PCR by using the pairs of primer 0-3 / 0-4 and 0-5 / 0-6 respectively. The PCR products were subsequently cloned into pTC*Verm* after digestion of the vector and of the inserts by *EcoRI* and *BamHI*. For pSU-*erm*::MiniTn*GBS1* construction, a fragment containing both IRs and *gbs0410* (the transposase gene of Tn*GBS1.1*) was amplified from Tn*GBS1.1* circular forms and by using the oligonucleotide pair 0-1 / 0-2. The *PstI*-*BamHI* amplicon was cloned into the pSU18-*erm* plasmid. pSU18-*erm* was obtained by cloning the Tn1545 erythromycin resistance gene [54] into the pACYCY184 derivative pSU18 [55]. pSU-*erm*::MiniTn*GBS2* was constructed by cloning the *HindIII* - *SalI* DNA fragment from pMUTIN::MiniTn*GBS2* [14] and encompassing the transposase gene and the two IRs into the pSU18-*erm* vector. The resulting plasmids are unable to replicate in *S. agalactiae*. The constructed plasmids were verified by sequencing and introduced into *S. agalactiae* strains by electroporation using electrocompetent cells prepared as previously described [56].

Strain GMP205 is a derivative of *S. agalactiae* strains A909 RF carrying a Tn*GBS1.1* with an inserted copy of pSU-*erm*::MiniTn*GBS2* in the *gbs0398* - *gbs0399* intergenic region. The insertion of the pSU derivative was obtained through a mating experiment between a NEM316 derivative carrying a pSU-*erm*::MiniTn*GBS2* inserted into the core genome and strain BM110. Tn*GBS1*::*erm* was subsequently transferred by conjugation to strain A909RF and the insertion of the ICE into the chromosome was obtained after several passages on TH plates supplemented with erythromycin. The insert position of Tn*GBS1.1* in strain GMP205 upstream *rpsJ* was identified by direct sequencing on chromosomal DNA.

**Bacterial conjugations.** The donor and the recipient strains were mixed together on a 3 mm hydrophobic edge membrane filter (Millipore, Billerica, MA) with a donor-to-recipient ratio of 1:1. Plates were incubated overnight at 37°C. The bacterial cells were resuspended in TH broth and spread on TH agar plates containing appropriate antibiotics. To quantify conjugative transfer efficiency, dilutions were plated on TH media containing antibiotics for the selection of transposon-containing donor cells. The

frequency of transfer was determined by dividing the number of TCs by the number of donor cells. All conjugation assays were performed in triplicates.

**Quantitative PCR.** Circular forms of *TnGBS1.1* and *TnGBS2.1* were quantified by comparing the amount of circular ICE DNA (primer pairs O-11 / O-12 and O-19 / O-20 amplifying the coupling sequence of *TnGBS1.1* and *TnGBS2.1* respectively) with the amount of circular and integrated forms of the ICE (primers O-13 / O-14 and O-21 / O-22 located in the transposase gene of *TnGBS1.1* and *TnGBS2.1* respectively). The relative quantity of specific DNA was normalized by the relative amplification of the two primer pairs using pSU::Mini*TnGBS1* and pMutin::Mini*TnGBS2* as a template in which these two regions are in equal amounts and by the using the *polA* gene as control (primers O-17 / O-18). DNA fragments were amplified with the ABI PRISM 7900 SDS system and SYBR Green PCR kits (Applied Biosystems, Foster City, USA). All measurements were performed at least in duplicate.

**Phylogenetic Analysis** For phylogenetic reconstruction all protein sequences, except the *TnGBS* sequences obtained in this work, were retrieved by BLASTP from the NCBI nr protein database using proteins of the *TnGBS1.1* or *TnGBS2.1* prototype as query. The different orthologous proteins were aligned using the MUSCLE algorithm [57] and blocks of poorly aligned positions and divergent regions of the alignment were removed using the Gblock tool (Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments). Maximum likelihood (ML) using PyML on the phylogeny.fr platform [58] was used to infer phylogenetic relationships. Prior to ML analysis, a protein substitution model for each dataset was selected using Prottest3 and the Akaike information criterion [59]. ML bootstrap support was determined using 100 bootstrap replicates.

## Acknowledgments

We wish to thank Mathieu Brochet, Patrick Trieu-Cuot, Isabelle Rosinski-Chupin, Maria-José Lopez-Sanchez and Laura Gomez for their interest and critical comments and Carmen Buchrieser for critical reading of the manuscript. This work was supported by the French National Research Agency (grants ANR-08-GENM-027-001 and 2010-PATH-004-02) and by the Fondation pour la Recherche Médicale (FRM).

## References

1. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS (2011) PFAST: a fast phage search tool. *Nucleic Acids Research* 39: W347-352.
2. Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, et al. (2008) The Whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. *DNA Res* 15: 185-199.
3. Davies J, Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* 74: 417-433.
4. Burrus V, Pavlovic G, Decaris B, Guedon G (2002) Conjugative transposons: the tip of the iceberg. *Molecular Microbiology* 46: 601-610.
5. Guglielmini J, Quintais L, Garcillan-Barcia MP, de la Cruz F, Rocha EP (2011) The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet* 7: e1002222.
6. Lederberg J, Tatum EL (1946) Gene recombination in *Escherichia coli*. *Nature* 158: 558.
7. Brochet M, Rusniok C, Couve E, Dramsi S, Poyart C, et al. (2008) Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A* 105: 15961-15966.
8. Manson JM, Hancock LE, Gilmore MS (2010) Mechanism of chromosomal transfer of *Enterococcus faecalis* pathogenicity island, capsule, antimicrobial resistance, and other traits. *Proc Natl Acad Sci U S A* 107: 12269-12274.
9. Schubert S, Darlu P, Clermont O, Wieser A, Magistro G, et al. (2009) Role of intraspecies recombination in the spread of pathogenicity islands within the *Escherichia coli* species. *PLoS Pathog* 5: e1000257.
10. Llosa M, Gomis-Ruth FX, Coll M, de la Cruz Fd F (2002) Bacterial conjugation: a two-step mechanism for DNA transport. *Molecular Microbiology* 45: 1-8.
11. Wallden K, Rivera-Calzada A, Waksman G (2010) Type IV secretion systems: versatility and diversity in function. *Cell Microbiol* 12: 1203-1212.
12. Cascales E, Christie PJ (2003) The versatile bacterial type IV secretion systems. *Nat Rev Microbiol* 1: 137-149.
13. Wozniak RA, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 8: 552-563.
14. Brochet M, Da Cunha V, Couve E, Rusniok C, Trieu-Cuot P, et al. (2009) Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. *Molecular Microbiology* 71: 948-959.
15. Clewell DB, Flannagan SE, Jaworski DD (1995) Unconstrained bacterial promiscuity: the Tn916-Tn1545 family of conjugative transposons. *Trends in Microbiology* 3: 229-236.

16. Poyart C, Celli J, Trieu-Cuot P (1995) Conjugative transposition of Tn916-related elements from *Enterococcus faecalis* to *Escherichia coli* and *Pseudomonas fluorescens*. *Antimicrobial Agents and Chemotherapy* 39: 500-506.
17. Marcinek H, Wirth R, Muscholl-Silberhorn A, Gauer M (1998) *Enterococcus faecalis* gene transfer under natural conditions in municipal sewage water treatment plants. *Applied and Environmental Microbiology* 64: 626-632.
18. Franken C, Haase G, Brandt C, Weber-Heynemann J, Martin S, et al. (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.
19. Brochet M, Couve E, Glaser P, Guedon G, Payot S (2008) Integrative conjugative elements and related elements are major contributors to the genome diversity of *Streptococcus agalactiae*. *Journal of Bacteriology* 190: 6913-6917.
20. Weaver KE, Kwong SM, Firth N, Francia MV (2009) The RepA\_N replicons of Gram-positive bacteria: a family of broadly distributed but narrow host range plasmids. *Plasmid* 61: 94-109.
21. Kwong SM, Skurray RA, Firth N (2004) *Staphylococcus aureus* multiresistance plasmid pSK41: analysis of the replication region, initiator protein binding and antisense RNA regulation. *Molecular Microbiology* 51: 497-509.
22. Francia MV, Weaver KE, Goicoechea P, Tille P, Clewell DB (2007) Characterization of an active partition system for the *Enterococcus faecalis* pheromone-responding plasmid pAD1. *Journal of Bacteriology* 189: 8546-8555.
23. Le Chatelier E, Janniere L, Ehrlich SD, Canceill D (2001) The RepE initiator is a double-stranded and single-stranded DNA-binding protein that forms an atypical open complex at the onset of replication of plasmid pAMbeta 1 from Gram-positive bacteria. *Journal of Biological Chemistry* 276: 10234-10246.
24. Bidnenko V, Ehrlich SD, Janniere L (1998) In vivo relations between pAMbeta1-encoded type I topoisomerase and plasmid replication. *Molecular Microbiology* 28: 1005-1016.
25. Tinsley E, Naqvi A, Bourgogne A, Koehler TM, Khan SA (2004) Isolation of a minireplicon of the virulence plasmid pX02 of *Bacillus anthracis* and characterization of the plasmid-encoded RepS replication protein. *Journal of Bacteriology* 186: 2717-2723.
26. Weaver KE, Clewell DB, An F (1993) Identification, characterization, and nucleotide sequence of a region of *Enterococcus faecalis* pheromone-responsive plasmid pAD1 capable of autonomous replication. *Journal of Bacteriology* 175: 1900-1909.
27. Liyo VS, Pratto F, de la Hoz AB, Ayora S, Alonso JC (2010) Plasmid pSM19035, a model to study stable maintenance in Firmicutes. *Plasmid* 64: 1-17.
28. Francia MV, Varsaki A, Garcillan-Barcia MP, Latorre A, Drainas C, et al. (2004) A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol Rev* 28: 79-100.
29. Tomita H, Ike Y (2005) Genetic analysis of transfer-related regions of the vancomycin resistance *Enterococcus* conjugative plasmid pHTbeta: identification of oriT and a putative relaxase gene. *Journal of Bacteriology* 187: 7727-7737.
30. Zheng B, Tomita H, Inoue T, Ike Y (2009) Isolation of VanB-type *Enterococcus faecalis* strains from nosocomial infections: first report of the isolation and identification of the pheromone-responsive plasmids pMG2200, Encoding VanB-type vancomycin resistance and a Bac41-type bacteriocin, and pMG2201, encoding erythromycin resistance and cytotoxin (Hly/Bac). *Antimicrobial Agents and Chemotherapy* 53: 735-747.
31. Poyart C, Trieu-Cuot P (1997) A broad-host-range mobilizable shuttle vector for the construction of transcriptional fusions to beta-galactosidase in gram-positive bacteria. *FEMS Microbiology Letters* 156: 193-198.
32. Chen Y, Staddon JH, Dunny GM (2007) Specificity determinants of conjugative DNA processing in the *Enterococcus faecalis* plasmid pCF10 and the *Lactococcus lactis* plasmid pRS01. *Molecular Microbiology* 63: 1549-1564.

33. Garcillan-Barcia MP, Francia MV, de la Cruz F (2009) The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev* 33: 657-687.
34. Roberts AP, Mullany P (2009) A modular master on the move: the Tn916 family of mobile genetic elements. *Trends in Microbiology* 17: 251-258.
35. Wozniak RA, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, et al. (2009) Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet* 5: e1000786.
36. Glaser P, Rusniok C, Buchrieser C, Chevalier F, Frangeul L, et al. (2002) Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol Microbiol* 45: 1499-1513.
37. Brady LJ, Maddocks SE, Larson MR, Forsgren N, Persson K, et al. (2010) The changing faces of Streptococcus antigen I/II polypeptide family adhesins. *Molecular Microbiology* 77: 276-286.
38. Maddocks SE, Wright CJ, Nobbs AH, Brittan JL, Franklin L, et al. (2011) *Streptococcus pyogenes* antigen I/II-family polypeptide AspA shows differential ligand-binding properties and mediates biofilm formation. *Molecular Microbiology* 81: 1034-1049.
39. Sorensen UB, Poulsen K, Ghezzi C, Margarit I, Kilian M (2010) Emergence and global dissemination of host-specific *Streptococcus agalactiae* clones. *MBio* 1.
40. Richards VP, Lang P, Bitar PD, Lefebvre T, Schukken YH, et al. (2011) Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*. *Infect Genet Evol* 11: 1263-1275.
41. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research* 34: D32-36.
42. Smyth DS, Robinson DA (2009) Integrative and sequence characteristics of a novel genetic element, ICE6013, in *Staphylococcus aureus*. *Journal of Bacteriology* 191: 5964-5975.
43. Han X, Ito T, Takeuchi F, Ma XX, Takasu M, et al. (2009) Identification of a novel variant of staphylococcal cassette chromosome mec, type II.5, and its truncated form by insertion of putative conjugative transposon Tn6012. *Antimicrobial Agents and Chemotherapy* 53: 2616-2619.
44. Kiss J, Szabo M, Olasz F (2003) Site-specific recombination by the DDE family member mobile element IS30 transposase. *Proc Natl Acad Sci U S A* 100: 15000-15005.
45. Kiss J, Olasz F (1999) Formation and transposition of the covalently closed IS30 circle: the relation between tandem dimers and monomeric circles. *Molecular Microbiology* 34: 37-52.
46. Ton-Hoang B, Betermier M, Polard P, Chandler M (1997) Assembly of a strong promoter following IS911 circularization and the role of circles in transposition. *EMBO Journal* 16: 3357-3371.
47. Campbell A (1981) Evolutionary significance of accessory DNA elements in bacteria. *Annu Rev Microbiol* 35: 55-83.
48. Wozniak RA, Waldor MK (2009) A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS Genet* 5: e1000439.
49. Grohmann E (2010) Autonomous plasmid-like replication of *Bacillus* ICEBs1: a general feature of integrative conjugative elements? *Molecular Microbiology* 75: 261-263.
50. Lee CA, Babic A, Grossman AD (2010) Autonomous plasmid-like replication of a conjugative transposon. *Molecular Microbiology* 75: 268-279.
51. Babic A, Berkmen MB, Lee CA, Grossman AD (2011) Efficient gene transfer in bacterial cell chains. *MBio* 2.
52. Kirchner J, Connolly CM, Sandmeyer SB (1995) Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. *Science* 267: 1488-1491.
53. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18: 821-829.



54. Trieu-Cuot P, Poyart-Salmeron C, Carlier C, Courvalin P (1990) Nucleotide sequence of the erythromycin resistance gene of the conjugative transposon Tn1545. *Nucleic Acids Research* 18: 3660.
55. Martinez E, Bartolome B, de la Cruz F (1988) pACYC184-derived cloning vectors containing the multiple cloning site and lacZ alpha reporter gene of pUC8/9 and pUC18/19 plasmids. *Gene* 68: 159-162.
56. Cruz-Rodz AL, Gilmore MS (1990) High efficiency introduction of plasmid DNA into glycine treated *Enterococcus faecalis* by electroporation. *Molecular and General Genetics* 224: 152-154.
57. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
58. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* 36: W465-469.
59. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104-2105.
60. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31: 3497-3500.
61. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14: 1188-1190.
62. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan MS, et al. (2003) Multilocus sequence typing system for group B *Streptococcus*. *J Clin Microbiol* 41: 2530-2536.
63. Haenni M, Saras E, Bertin S, Leblond P, Madec JY, et al. (2010) Diversity and mobility of integrative and conjugative elements in bovine isolates of *Streptococcus agalactiae*, *S. dysgalactiae subsp. dysgalactiae*, and *S. uberis*. *Applied and Environmental Microbiology* 76: 7957-7965.
64. Stalhammar-Carlemalm M, Stenberg L, Lindahl G (1993) Protein rib: a novel group B streptococcal cell surface protein that confers protective immunity and is expressed by most strains causing invasive infections. *J Exp Med* 177: 1593-1603.

## Figure legends

**Figure 1.** Distribution of ICEs of the Tn*GBS* family among streptococci. Streptococcal species are represented according to the phylogenetic tree inferred from comparison of the RecN protein sequences using the neighbor-joining method. Bar, 0.01 changes per amino acid position. Bootstrap values higher than 50 (from 100 replicates) are indicated. Numbers in parentheses indicate, for species with at least one ICE identified, the number of genomes containing at least one Tn*GBS* ICE versus the number of genomes sequenced available at the NCBI. The \* indicates that *S. dysgalactiae subsp. equisimilis* elements are non conjugative.

**Figure 2.** Distribution of conserved proteins among Tn*GBS*1-like ICEs and related conjugative plasmids. Tn*GBS*1.1 of strain NEM316 was used as a reference. A filled box indicates the presence of a gene expressing a similar protein. The intensity of the blue corresponds to the % of identities according to the indicated scale. For related plasmids the presence of a gene coding for a similar protein or a protein showing a similar domain is indicated by a grey box. Boxes marked with a white star correspond to genes with a frameshift mutation or a sequencing error. The gene encoding the DDE transposase and genes coding for protein belonging to the mobilization and DNA processing, T4SS, replication, and surface adhesion are respectively boxed in red, orange, green, yellow and purple. ICE accession numbers are indicated in Table 1. Plasmid accession numbers are: pMP118, NC\_007930; pAW63, NC\_010599; pFR55, NC\_010283; pHT $\beta$ , AB183714; pMG1, NC\_011364.

**Figure 3.** Distribution of conserved proteins among Tn*GBS*2-like ICEs and related conjugative plasmids and ICEs depending on a phage like integrase. Tn*GBS*2.1 of strain NEM316 was used as reference. A filled box indicates the presence of a gene expressing a similar protein. The intensity of the blue corresponds to the % of identities. For plasmids and ICEs the presence of a gene coding a similar protein or a protein showing a similar domain is indicated by a grey box. Boxes marked with a white star correspond to genes with a frame shift mutation or a sequencing error. The gene encoding the DDE transposase and genes coding for proteins belonging to the mobilization and DNA processing, T4SS, replication, and surface adhesion are respectively boxed in red, orange, green, yellow and purple. Tn*GBS*2 related ICE accession numbers are indicated in Table

1. Plasmid and ICE accession numbers are: pTEF2, NC\_004671.1; pCF10, AY855841.2; Sex factor; pIL6, HM021331.1; Tn5253, EU351020.1; ICESde3396, EU142041; 89K pathogenicity island of *S. suis*, EU589333.1; pBD-II.

**Figure 4.** Phylogeny of the TnGBS ICE transposases. Maximum likelihood (ML) using PyML on the phylogeny.fr platform [58] was used to infer phylogenetic relationships. ML bootstrap support was determined using 100 bootstrap replicates. Only bootstrap values higher than 50 are indicated. The sequence of the transposases of *L. crispatus*, *L. helveticus* and *B. coagulans* ISS were used as outgroup. The names correspond to the ICE name given in Table 1. Stars indicate TnGBS2 like ICE expressing a ParA homolog

**Figure 5.** Alignment of the IR-R and IR-L of TnGBS ICEs. IR Sequences were aligned by using CLUSTALW [60] and the Logo generated by using the WebLogo web site [61]. ICE names are according to Table 1. Conserved bases are indicated in blue and non conserved ones in red. Bases marked by a star are complementary in IR-L and IR-R.

**Figure 6.** Putative origin of replication of TnGBS1.1 and TnGBS2.1. A The putative origin of replication located with the *repA* homolog ORF of TnGBS2.1 was aligned with the origin of replication of *S. aureus* plasmid pSK41 and with similar sequences identified in TnSsang2.1 and TnS.oral2.1 *repA* homologs. The repeated sequences are indicated in four different colors. Non-conserved bases are indicated in grey. B. The putative origin of replication located downstream TnGBS1.1 *repE* homolog was aligned with validated or hypothetical origins of replication of Gram-positive plasmids. Conserved residues are in red and the conserved C that corresponds to the initiation of replication is marked by an arrow.

**Figure 7.** TnGBS1.1 *oriT* region. The predicted origin of transfer to TnGBS1.1 located upstream *gbs0380* was aligned with validated or hypothetical *oriT* sites of Gram-positive plasmids. The two inverted repeats are shown in green and blue respectively, the conserved sequence in bold.

**Figure 8.** TnGBS2.1 *oriT* region. The conserved parts of the predicted origin of transfer of selected TnGBS2-like ICEs located upstream *gbs1121* in TnGBS2.1 were aligned with the

validated or hypothetical *oriT* sites of Gram-positive plasmids and ICEs. The inverted repeat is indicated in blue and the conserved sequence that includes the nick site in bold. Demonstrated nick sites in pC221, pRS01 and pCF10 are indicated by a small arrow.

**Figure S1:** Diversity of Tn*GBS1* related ICE. The gene encoding the DDE transposase is colored in red. The genes coding for proteins belonging to the mobilization and DNA processing, T4SS formation, replication, and surface adhesion are respectively colored in orange, green, yellow and purple. The genes coding proteins with a VirD4, VirB4 and VirB6 domain are indicated in bright green. The gene encoding the relaxase and a protein with a ClpA domain are colored in dark orange, and brown respectively.

**Figure S2:** Diversity of Tn*GBS2* related ICE. The gene encoding the DDE transposase is colored in red. The genes coding for protein belonging to the mobilization and DNA processing, T4SS formation, replication, and surface adhesion are respectively colored in orange, green, yellow and purple. The genes coding proteins with VirD4, VirB4 and VirB6 domain are indicated in bright green. The gene encoding the relaxase is colored in dark orange.

**Figure S3:** Tn*GBS1* and the integrative plasmid pSU18::*erm*-miniTn*GBS1*. A. circular form and B. inserted form of the non replicative plasmid pSU18::*erm*-miniTn*GBS1*. C schematic representation of the circular and D. inserted copy of Tn*GBS1*.1. Genes, the pACYC184 origin of replication and inverted repeats (IRs) are indicated by arrows. The DNA fragment in pSU18::*erm*-miniTn*GBS1* originating from Tn*GBS1*.1 is in orange. Oligonucleotides used for detection and quantification of the circular form by Q-PCR are indicated by small arrows. O-11, O-12 to quantify the circular form and O-13, O-14 to quantify *gb0410*.

**Figure S4.** DNA sequences of the insertion sites of Tn*GBS1* and pSU18*erm*::MiniTn*GBS1*. *gbs* numbers correspond to the nomenclature of the orthologous genes in strains NEM316. Numbers of identified insertions per site are indicated on the right for 16 BM110 TCs and 61 and 23 pSU18*erm*::MiniTn*GBS1* transformants in NEM316 and in BM110 respectively. Blue letters represent the right extremity of the transposable

element (IR-R); Underlined letters, the sequence duplicated after insertion (DR); Red letters, putative -35 and -10 elements; Purple letters, putative extended -10 region.

**Figure S5:** Phylogeny of Tn*GBS1* related conjugative module. Maximum likelihood (ML) using PyML on the phylogeny.fr platform [58] was used to infer phylogenetic relationships. A relaxase, B VirB4 and C VirD4 proteins.

**Figure S6:** Phylogeny of Tn*GBS2* related conjugative module. Maximum likelihood (ML) using PyML on the phylogeny.fr platform [58] was used to infer phylogenetic relationships. A relaxase, B VirB4 and C VirD4 proteins

**Table 1.** Transfer efficiencies of TnGBS1.1 and TnGBS2.1 according to different recipient strains

Recipient species	<sup>1</sup> Transfer Efficiency	
	TnGBS1	TnGBS2
<i>S. agalactiae</i> BM110	7.2 .10 <sup>-5</sup> <sup>2</sup> (1.4)	2.4 .10 <sup>-5</sup> ( 2.5)
<i>S. gallolyticus</i> UCN34	<sup>3</sup> 0.5 . 10 <sup>-9</sup>	7 .10 <sup>-8</sup> ( 5.9)
<i>S. thermophilus</i> LMD9	<sup>4</sup> ND	ND
<i>E. faecalis</i> JH2-2	ND	ND

<sup>1</sup>Transfer efficiency corresponds to the number of TCs per recipient bacteria. <sup>2</sup>the standard deviation is given in parenthesis; <sup>3</sup>this value correspond to a single TC per conjugation, <sup>4</sup>ND means that no TCs were obtained

Table 2: List of TnGBS related ICEs

ICE name <sup>1</sup>	Strain	GC% ICE	GC% host	<sup>3</sup> insertion site	length kb	accession number
<b>TnGBS1 - like</b>						
TnGBS1.1 a to c	<i>S. agalactiae</i> NEM316	37.6%	35.6%	<i>gbs0411-gbs0741-gbs0968</i>	47	NC_004368
<sup>2</sup> TnGBS1.2	<i>S. agalactiae</i> ATCC 13813	-	34.9% <sup>4</sup>	<i>galE</i>	-	AEQQ01000019
TnGBS1.3	<i>S. agalactiae</i> FSL S3-026	38.5%	36.1%	<i>rpsU</i>	46	AEXT01000007
TnGBS1.4 a to b	<i>S. agalactiae</i> FSL S3-026	38.5%	36.1%	<i>rpoB, polC</i>	50	AEXT01000002, 8
TnSgallo1.1	<i>Streptococcus gallolyticus</i> ATCC 43143	38%	37.5%	<i>tRNA-Cys</i>	51	AP012053
<b>TnSpneum1.1</b>						
TnSpneum1.1	<i>S. pneumoniae</i> GA17328	37.1%	39.7% <sup>4</sup>	<i>NI</i>	46	AGPH01000017
TnSspX1.1	<i>Streptococcus</i> sp. 2_1_36FAA	40.8%	40.5%	<i>glmS</i>	40	NZ_GG704942
TnSspora1.1	<i>S. taxon oral</i> sp taxon 071 str. 73H25AP	36.9%	41.5% <sup>4</sup>	<i>rpsO</i>	46	NZ_AEEP01000006
<b><sup>2</sup>TnSsang1.1</b>						
<sup>2</sup> TnSsang1.1	<i>S. sanguinis</i> SK1058	-	-	<i>pyrG</i>	-	AFBF01000017
TnSang1.1	<i>S. anginosus</i> F0211	34.6%	37.6% <sup>4</sup>	<i>murF</i>	52	AECT01000063
<sup>2</sup> TnSinf1.1	<i>S. infantis</i> SK970	-	-	<i>rpsL</i>	-	AFUT01000004-08
TnSinf1.2	<i>S. infantis</i> X	37.1%	39.0% <sup>4</sup>	<i>rpoE</i>	44	AFUQ01000001
TnSpast1.1	<i>S. pasteurianus</i> ATCC 43144	-	-	<i>rpsP</i>	-	NC_015600
TnSuri1.1	<i>S. urinalis</i> 2285-97	30%	34.1%	<i>rodA</i>	44	AEUZ01000004
<sup>2</sup> TnSsob1.1	<i>S. sobrinus</i> TCI-54	-	-	<i>eno</i>	-	AGGH01000002
<sup>2</sup> TnSsob1.2	<i>S. sobrinus</i> TCI-89	-	-	<i>rpmV</i>	-	AGGN01000040
<sup>2</sup> TnSsob1.3	<i>S. sobrinus</i> TCI-118	-	-	<i>rpmV</i>	-	AGHN01000009
<sup>2</sup> TnSsob1.4	<i>S. sobrinus</i> TCI-342	-	-	<i>dltA</i>	-	AGGT01000092
<sup>2</sup> TnSsob1.5	<i>S. sobrinus</i> TCI-384	-	-	<i>rpmV</i>	-	AGHI01000004
<sup>2</sup> TnSsob1.6	<i>S. sobrinus</i> TCI-61	-	-	<i>NI</i>	-	AGGJ01000234
<sup>2</sup> TnLsal1.1	<i>L. salivarius</i> ATCC11741	-	-	<i>NI</i>	-	ACGT01000030
<b>TnGBS2 - like</b>						
TnGBS2.1	<i>S. agalactiae</i> NEM316	38.0%	35.6%	<i>xpt</i>	34	NC_004368
<sup>2</sup> TnGBS2.2	<i>S. agalactiae</i> ATCC 13813	-	34.9% <sup>4</sup>	<i>glyQ</i>	-	AEQQ01000097-98

<b>TnGBS2.3</b>	<i>S. agalactiae</i> Wc3	38.1%	xx	<i>fbsA</i>	33	This study
<b>TnGBS2.4</b>	<i>S. agalactiae</i> CZ183	38.3%	xx	<i>guaC</i>	29	This study
<b>TnGBS2.5</b>	<i>S. agalactiae</i> 1122	34.4%	xx	<i>rpmH</i>	38	This study
<b>TnGBS2.6</b>	<i>S. agalactiae</i> 2584	34.5%	xx	<i>rplJ</i>	45	This study
<b>TnGBS2.7</b>	<i>S. agalactiae</i> FSL S3-026	38%	36.1%	<i>guaC</i>	34	AEXT01000007
<b>TnSinf2.1</b>	<i>S. infantis</i> SK970	39.1%		<i>rpmV</i>	32	AFUT01000008
<b><sup>2</sup>TnSinf2a2.1</b>	<i>S. infantarius</i> ATCC BAA-102		37.6%	<i>rpmH</i>		NZ_ABJK02000018
<b>TnSgallo2.1</b>	<i>S. gallolyticus</i> UCN34	36.7%	37.6%	<i>rpmB</i>	41	NC_013798
<b>TnSgallo2.2</b>	<i>S. gallolyticus</i> ATCC BAA-2069	38.8%	37.8%	<i>rpmH</i>	36	NC_015215
<b>TnSgallo2.3</b>	<i>S. gallolyticus</i> ATCC BAA-2069	39.2%	37.8%	$\S$	30	NC_015215
<b>TnSoral2.1</b>	<i>S. oralis</i> ATCC 35037	39.3%	40.3% <sup>4</sup>	NI	29	NZ_ADMV01000022
<b>TnSoral2.2</b>	<i>S. oralis</i> SK255	39.6%	41.5% <sup>4</sup>	HMPREF9968_1062	29	AFNM01000055
<b>TnSsporal2.1</b>	<i>S. taxon oral</i> sp taxon 071 str. 73H25AP	39.1%	41.5% <sup>4</sup>	<i>glyQ</i>	31	NZ_AEEP01000011
<b>TnSspX2.1</b>	<i>Streptococcus</i> sp. 2_1_36FAA	41.6%	40.5%	<i>rpsF</i>	-	NZ_GG704942
<b>TnSmit2.1</b>	<i>S. mitis</i> SK597	37.6%	40.2% <sup>4</sup>	<i>pyrF</i>	30	NZ_AEDV01000027
<b><sup>2</sup>TnSmit2.2</b>	<i>S. mitis</i> SK1073	-	40.6% <sup>4</sup>	<i>pyrF</i>	-	AFQT01000042
<b>TnSsang2.1</b>	<i>S. sanguinis</i> ATCC 49296	39.3%	41.6% <sup>4</sup>	<i>rpmB</i>	32	NZ_AEPO01000010
<b>TnSsang2.2</b>	<i>S. sanguinis</i> SK1059	38.9%	42.5% <sup>4</sup>	<i>rpsB</i>	25	AFFM01000023
<b><sup>2</sup>TnSsang2.3</b>	<i>S. sanguinis</i> SK49	-	-	<i>relA2</i>	-	AFFO01000015-16
<b>TnSsang2.4</b>	<i>S. sanguinis</i> SK72	41.7%	42.3% <sup>4</sup>	<i>pyrG</i>	25	AEXV01000011
<b><sup>2</sup>TnSsang2.5</b>	<i>S. sanguinis</i> SK29667	-	-	<i>rpsB</i>	-	AFQA01000023-24
<b>TnSsang2.6</b>	<i>S. sanguinis</i> SK340	39.0%	42.0% <sup>4</sup>	<i>rpsB</i>	25	AFQB01000011
<b>TnSsang2.7</b>	<i>S. sanguinis</i> SK1058	-	-	<i>pyrG</i>	-	AFBF01000017
<b>TnSsui2.1</b>	<i>S. suis</i> D12	40.8%	41.3%	<i>rplU</i>	33	CP002644
<b>TnScons2.1</b>	<i>S. constellatus</i> SK1060	38.2%	38.1%	<i>rpmH</i>	25	NZ_AFUP01000009
<b><sup>2</sup>TnSict2.1</b>	<i>S. ictaluri</i> 707-05	-	39.2% <sup>4</sup>	ABC transporter	-	NZ_AEUX02000006
<b><sup>2</sup>TnSict2.2</b>	<i>S. ictaluri</i> 707-05	-	39.2% <sup>4</sup>	<i>rimP</i>	-	NZ_AEUX02000007
<b>TnSang2.1</b>	<i>S. anginosus</i> SK52	37.8%	39.1% <sup>4</sup>	NI	33	NZ_AFIM01000003
<b>TnSzoo2.1</b>	<i>S. equi</i> subsp. <i>zooepidemicus</i>	39.2%	41.6%	<i>rpmB</i>	32	CP002904



ATCC 35246							
<b>TnSsob2.1</b>	<i>S. sobrinus</i> TCI-396	41.1%	43.0% <sup>4</sup>	<i>leuA</i>	27	AGHL01000006	
<b><sup>2</sup>TnSsob2.2</b>	<i>S. sobrinus</i> TCI-345	-	-	<i>cspR</i>	-	AGGU010000185	
<b><sup>2</sup>TnSsob2.3</b>	<i>S. sobrinus</i> TCI-366	-	-	<i>leuA</i>	-	AGHB010000008	
<b><sup>2</sup>TnSsob2.4</b>	<i>S. sobrinus</i> TCI-377	-	-	NI	-	AGHG010000295	
<b><sup>5</sup>TnSdys2.1</b>	<i>S. dysgalactiae</i> subsp. <i>equisimilis</i> GGS_124	35.6%	39.6%	<i>SDEG_1257 ABC transporter</i>	6	NC_012891.1	
<b><sup>5</sup>TnSdys2.2</b>	<i>S. dysgalactiae</i> subsp. <i>equisimilis</i> ATCC 12394	35.7%	39.5%	<i>SDE12394_06580</i>	6	AFUL010000001	

<sup>1</sup> The name includes the initial of the species followed by 1 or 2 indicating whether the ICE is related to TnGBS1 or TnGBS2 respectively and by an order number

<sup>2</sup> The TnGBS sequence is incomplete and split in several contigs.

<sup>3</sup> The gene name corresponds to the promoter upstream which the ICE is inserted.

<sup>4</sup> The GC% has been estimated using the draft genome.

<sup>5</sup> These two elements correspond to simple transposon.

NI: not identifiable

**Table 3.** sequences of the coupling sequence in BM110 transconjugants

<sup>1</sup> Target gene	<sup>2,3</sup> TnGBS2	<sup>2</sup> TnGBS1
<sup>4</sup> <i>gbs1117</i>	108 (CATTTAAA)	171 (CCATTTAAA)
<i>nrdF</i>	2 (ATTTTTCAA)	3 (TTTTTCAA)
<i>gbs1262</i>	1 (TTATAACGC)	
Unknwon		1 (GATTTTTAAA)
<sup>5</sup> <i>rpsJ</i>		151 (AAATAAA)

<sup>1</sup>Gene names correspond to NEM316 nomenclature. <sup>2</sup>sequence of the coupling sequence is given in parenthesis. <sup>3</sup>two independent TCs were analyzed for TnGBS2.1; <sup>4</sup>*gbs1117* corresponds to the insertion site of TnGBS2.1 in the donor strain. <sup>5</sup>*rpsJ* corresponds to the insertion site of TnGBS1.1 in the donor strain.

**Table 4.** Quantification of the ICE circular form and transfer efficiency

Donor	Recipient	<sup>1</sup> Transfer efficiency	Circular form / TnGBS	Efficiency/ Circular form
<b>TnGBS2</b>				
GMP201	BM110	2. 10 <sup>-5</sup> (0.7)	8 10 <sup>-4</sup> (0.3)	0.025
<sup>2</sup> TC BM110	BM110 RF	3.1 10 <sup>-3</sup> (2.5)	2.6 10 <sup>-1</sup> (0.78)	0.08
TC BM110p	BM110 RF	3.5 10 <sup>-5</sup> (1.3)	<sup>4</sup> ND	-
<b>TnGBS1</b>				
GMP205	BM110	7.2 10 <sup>-5</sup> (1.7)	6.9 10 <sup>-2</sup> (2.4)	1 10 <sup>-3</sup>
<sup>2</sup> TC BM110	BM110 RF	1.6 10 <sup>-1</sup> (0.7)	1.1 (0.1)	0.145
<sup>3</sup> TC BM110p	BM110 RF	9.1 10 <sup>-5</sup>	ND	-

<sup>1</sup>Number of TCs per donor cell; <sup>2</sup>a TC directly after conjugation; <sup>3</sup>a TC after seven passages; <sup>4</sup>not determined

**Table 5:** TnGBS1.1 and TnGBS2.1 *oriT* site: mobilization efficiency of pTCV-erm derivatives

Donor strain and Plasmid	Recipient strain	
	<i>S. agalactiae</i> BM110	<i>E. faecalis</i> JH2-2
<i>NEM316</i> pTCV-erm $\Omega$ <i>oriT</i> <sub>TnGBS1</sub>	1.7 .10 <sup>-5</sup> (0.81)	1.5 .10 <sup>-6</sup>
<i>NEM316</i> pTCV-erm $\Omega$ <i>oriT</i> <sub>TnGBS2</sub>	2.2 .10 <sup>-6</sup> (0.3)	6.1.10 <sup>-6</sup>
<i>NEM316</i> pTCV-erm	7.2 .10 <sup>-9</sup> (2.6)	ND

ND means that no tranconjugant were obtained, with the detection limit of 10<sup>-10</sup> under the experimental conditions we have used.

**Table S1.** Bacterial strains and plasmids used for cloning and conjugation assays.

Strains and plasmids	Relevant genotype and resistance profile <sup>a</sup> or characteristics	Reference or source
<i>E. coli</i>		
DH5a		Invitrogen
<i>S. agalactiae</i>		
NEM316	Isolated from neonate blood culture	[36]
WC3/UK21	Blood, EOD	[62]
1122	Bovine origin	[63]
2584	Bovine origin	[63]
CZ183	Bovine origin	[39]
GMP201	Erm; NEM316-Tn <i>GBS2</i> ::erm	[14]
GMP205	Erm; A909RF-Tn <i>GBS1</i> ::erm	This work
BM110	Tet; Isolated from neonate blood culture	[64]
BM110 RF	BM110 spontaneous mutant for Rif and Fus	This work
A909RF	A909 spontaneous mutant for Rif and Fus	[36]
Plasmids		
pSU18	Cm, pACYC184 derivative	[55]
pS18erm	Cm, ery	This work
pTCVerm	Erm, Km; <i>oriR</i> pACYC184, <i>oriRpAMB1</i>	[31]

<sup>a</sup> Amp, Erm, Fus, Rif, Tet, are resistance to ampicillin, erythromycin, fusidic acid, rifampicin, and tetracycline, respectively.

Table S2: Oligonucleotides used in this study

Name	Sequence 5' => 3'	Application
O-1	AAACTGCAGGGCTCCCTTCTTCATGAGATT	pSU18::miniT
O-2	CGCGGATCCGATTGAGCGAAAGAGGTAGGG	
O-3	CGGAATCCGGTAAACTCTATGTTTCAGATACAATTG	Cloning TnGBS1.1 <i>oriT</i> region
O-4	CGGGATCCCGATGGTCCGATGTCAATTC	Cloning TnGBS1.1 <i>oriT</i> region
O-5	CTGAAAGAATTCTAATGACCTACTTATTTGAG	Cloning TnGBS2.1 <i>oriT</i> region
O-6	ACACAGGATCCCTTATTAAGTATTTGACAGTC	Cloning TnGBS2.1 <i>oriT</i> region
O-7	GCTTAGTTCTAATGATCCTTATA	Insertion upstream <i>rpsJ</i>
O-8	GATGTCTACCAACGTTTGTG	Insertion upstream <i>rpsJ</i>
O-9	CGTTTATGGACTAACTTATTTGT	Insertion upstream <i>gbs0411</i>
O-10	GTTTTTCCCAATCCATCAACAA	Insertion upstream <i>gbs0411</i>
O-11	ACTGTCAAGGGCTGAGCAAG	Circular form of TnGBS1.1 and Q-PCR
O-12	TTTCTCCGTTCTCCCCATAA	Circular form of TnGBS1.1 and Q-PCR
O-13	GGGAGCTGAGACGATGAGTC	Q-PCR, quantification <i>gbs0410</i>
O-14	CCGACCAAGTGTCTGGAGAT	Q-PCR, quantification <i>gbs0410</i>
O-17	GGAACATTATGAAGCCGATGA	Q-PCR, quantification <i>polA</i>
O-18	TCAAACCTCAGCAACTCCCTTT	Q-PCR, quantification <i>polA</i>
O-19	CGATTTTCTCGTTCTCTTAAGCTAT	Q-PCR, quantification circular form TnGBS2.1
O-20	CGTTCTCTTAGGCGACAAGG	Q-PCR, quantification circular form TnGBS2.1
O-21	TGAGGATAGTGGTTCACGACA	Q-PCR, quantification <i>gbs1118</i>
O-22	AATCCTGCTGGAGCGTATTG	Q-PCR, quantification <i>gbs1118</i>

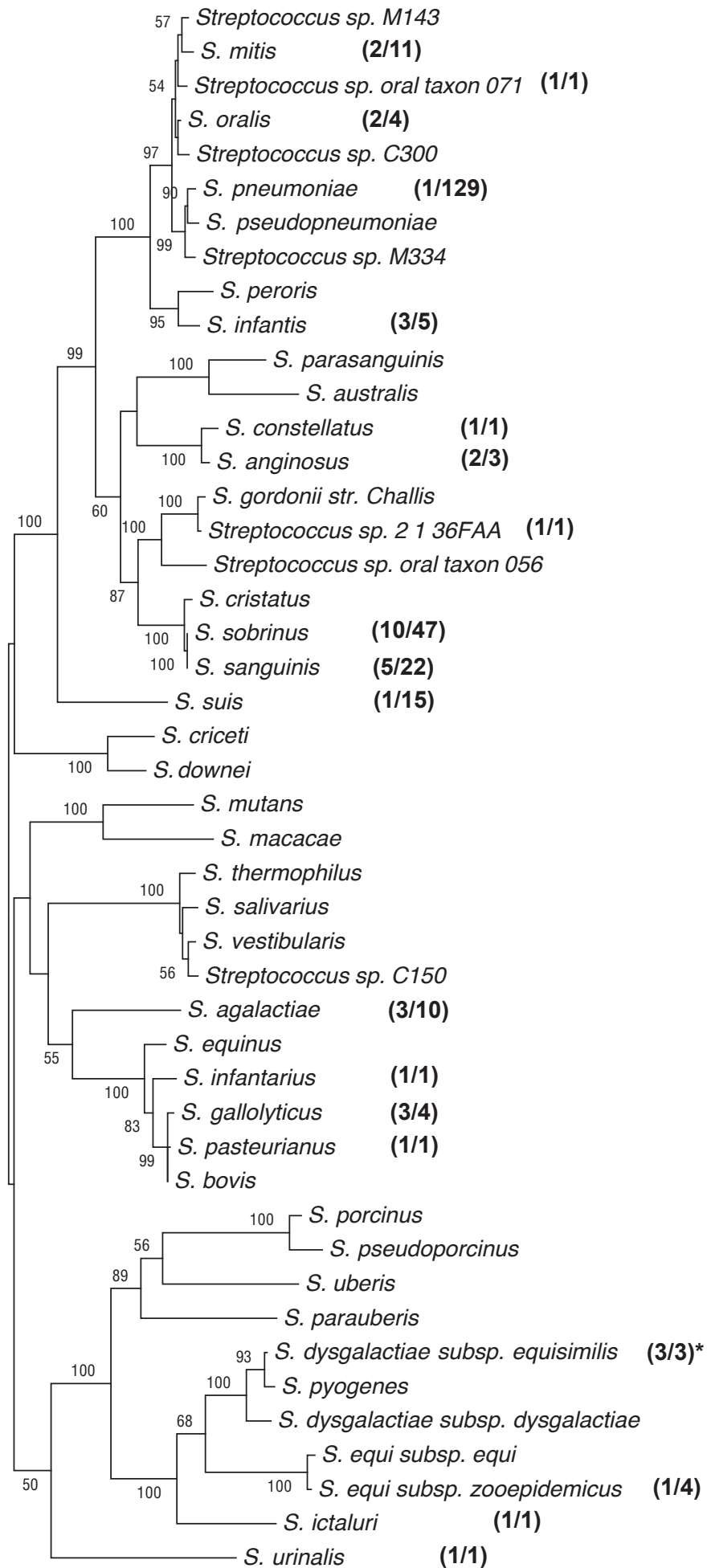


Figure 1

0.05

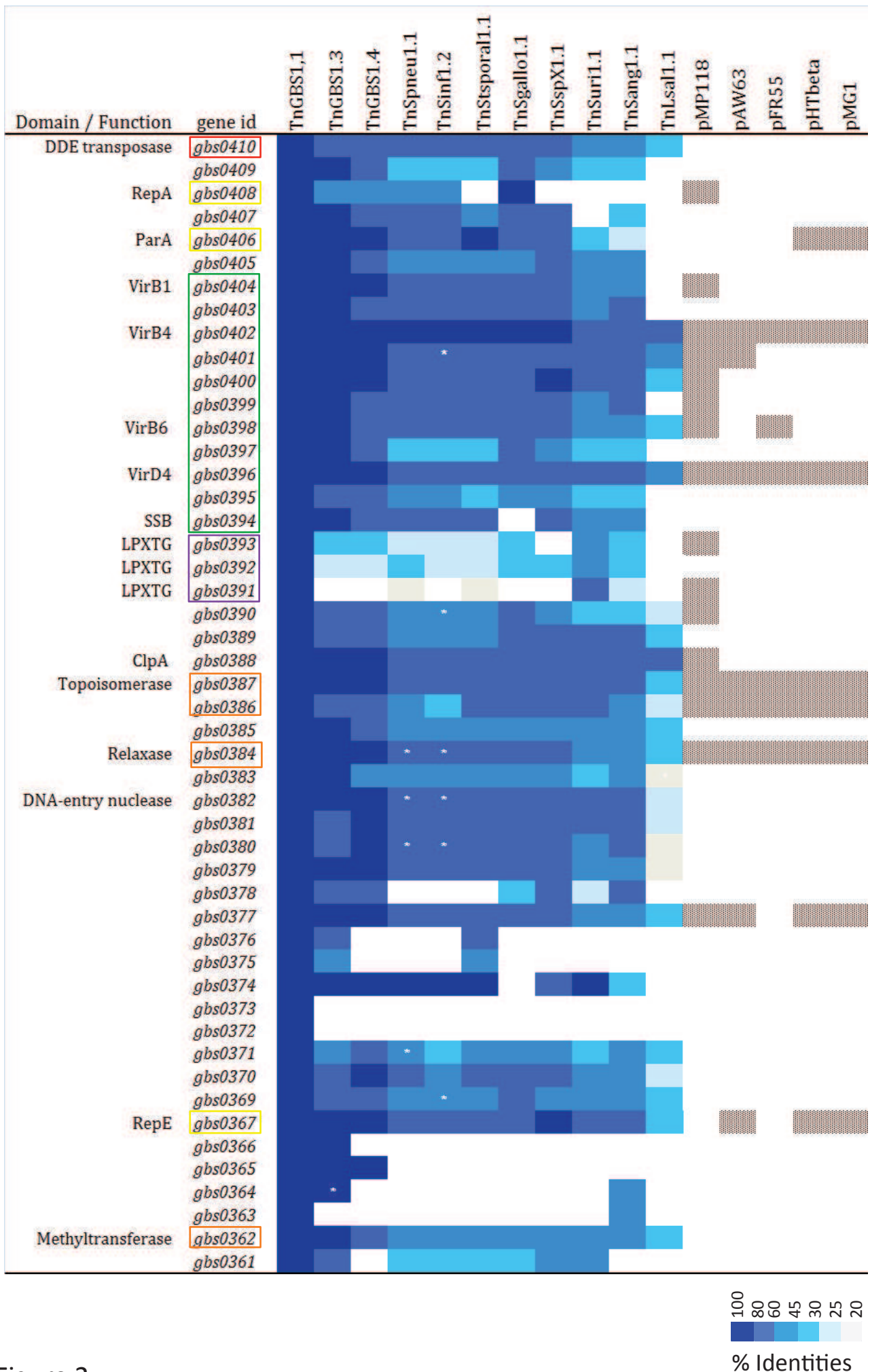


Figure 2



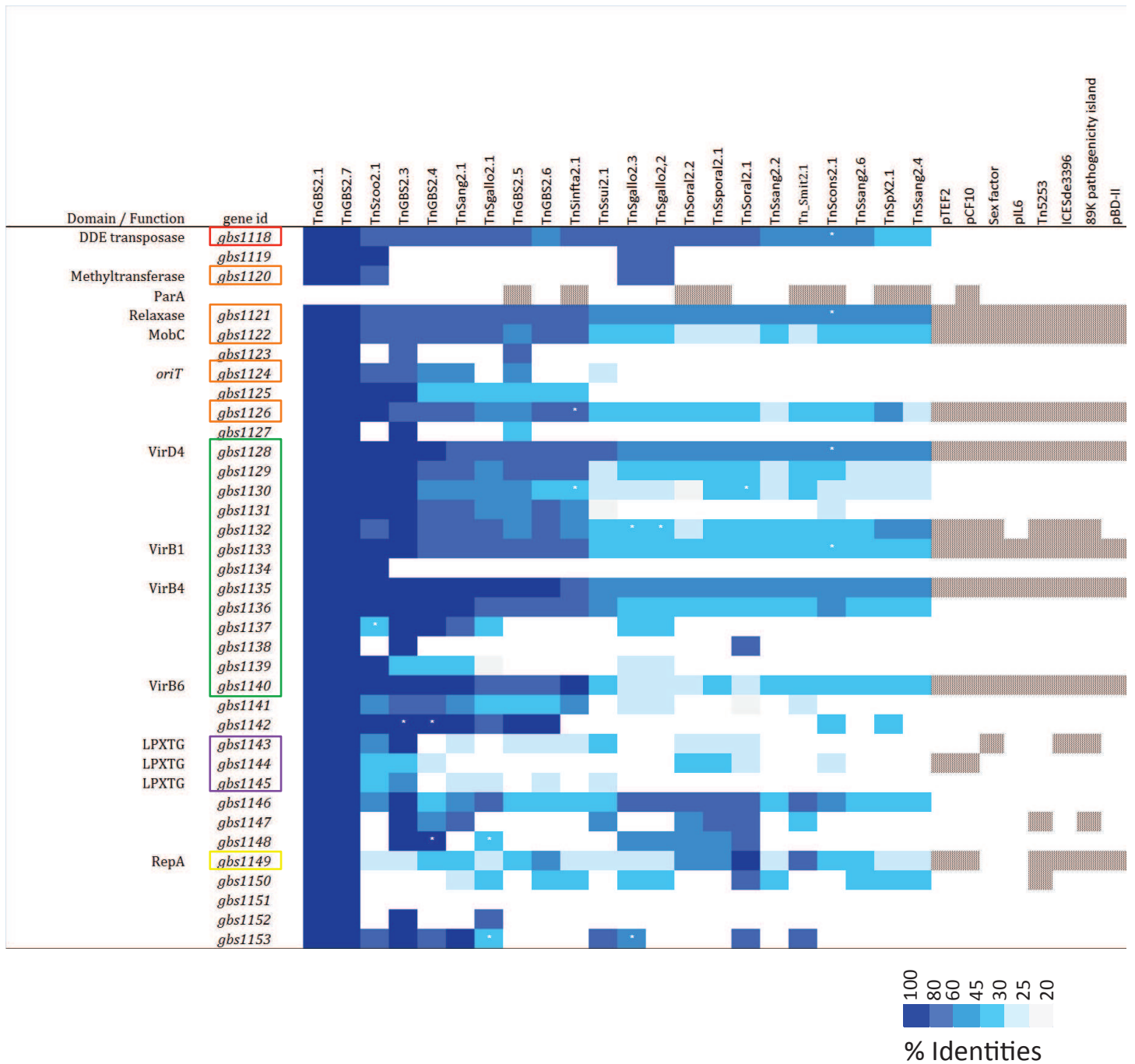


Figure 3

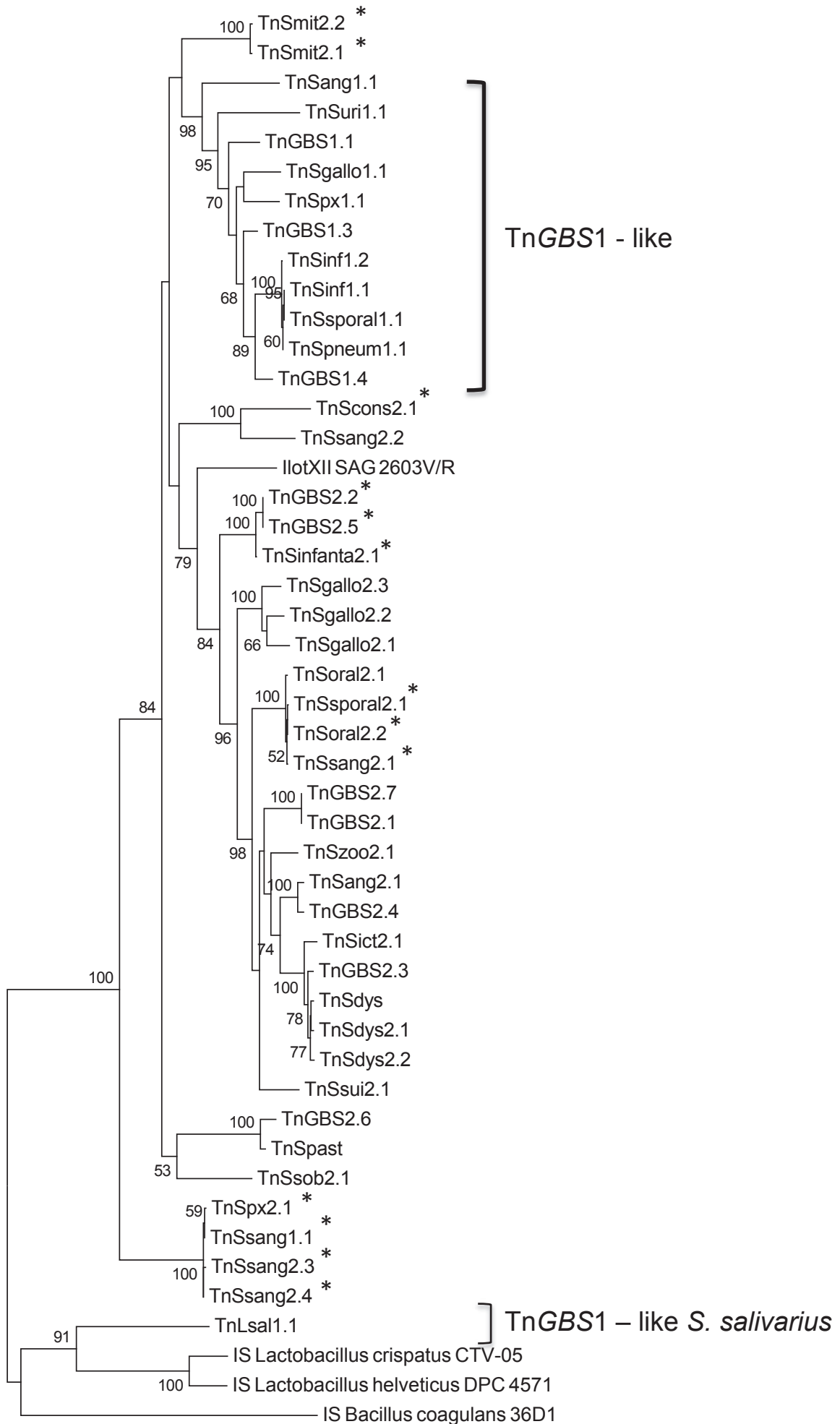

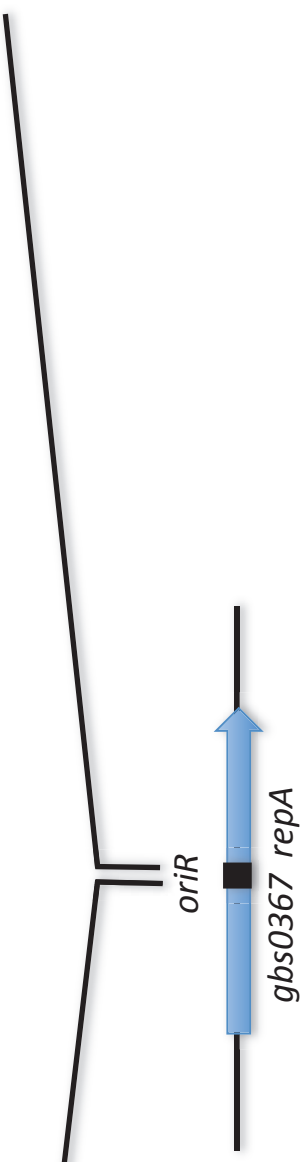


Figure 4 



A

	Rep box 1	Rep box 2	Rep box 3	Rep box 4
<b>pSK41</b>	GACAA <b>GGAA</b> TGT <b>CAAAA</b> GGACACGGTCCAGAAA	TGTCAAAA	GGACACGTTAG-AAATGTCAAAA	GGACACGTCTA <b>GAATGTCAAAA</b> GGACAC
<b>TnSsang2.1</b>	ACGGCGGAAATGT <b>CAAAA</b> CGACACTTCCAAAAA	-----	ATCAAGGAAATGTTAGAACGACATGT <b>CAAAAA</b> ATTCG <b>AAAA</b> CGACAC	
<b>TnSoral2.1</b>	ACAGCGGAAATGT <b>CAAAA</b> CGACACTTCCAAAAA	-----	ATCAAGGAAATGTCAAAA	TAA <b>CA</b> TCT <b>CAAAAA</b> ATGT <b>CGAAAA</b> CGACAC
<b>TnGBS2.2</b>	GTAGCGGAAAGTGT <b>CAAAA</b> CAACACTTCCAGAAA	-----	ATCAAG-AAATGTCAAAA	TGACACGT <b>CAAAAA</b> ATGT <b>CAAAA</b> CGACAC



B

<b>pX02</b>	TTTTAA	TTGTC	CACTC	TGCCA	AATAC	ATAG--	TATAT	CTA	CGATAC	
<b>pAW63</b>	TTTTAA	TTGTC	CACTC	TGCCA	AATAC	ATAA	TTATAT	CTT	CGATAC	
<b>pBT9727</b>	TTTTAA	TTGTC	CACTG	TGCCA	AATAC	ATAA	TTATAT	CTA	CGATAC	
<b>PSM19035</b>	AATAAAA	CCCGCA	CTA	TGCCA	TTAC	AT--	TTATAT	CTA	TGATAC	
<b>PIP501</b>	AATAAAA	CCCGCA	CTA	TGCCA	TTAC	AT--	TTATAT	CTA	TGATAC	
<b>pAMβ</b>	AATAAAA	CCCGCA	CTA	TGCCA	TTAC	AT--	TTATAT	CTA	TGATAC	
<b>TnSgal1o1.1</b>	AATAAAA	CCCGCA	AACTG	CC	TTAC	AT--	TTATAT	CTA	TGATAA	
<b>TnSpneum1.1</b>	TTAAAA	ACCCG	TAA	ACTG	CC--	TTAC	AT--	TTATAT	CTA	CGATTA
<b>TnGBS1.3</b>	GATAAAA	CCCGCA	AACTG	CC	TTAC	AT--	TTATAT	CT	TGATAC	
<b>TnGBS1.1</b>	GATAAAA	CCCGCA	AACTG	CC	TTAC	AT--	TTATAT	CT	TGATAC	

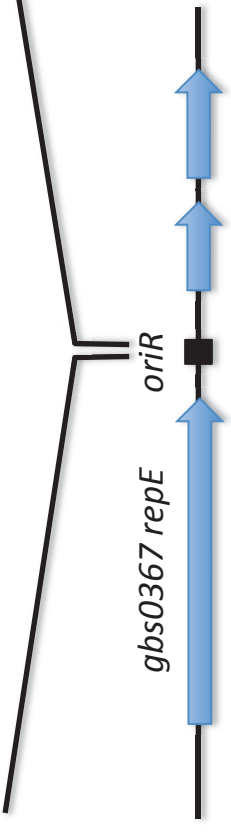


Figure 6

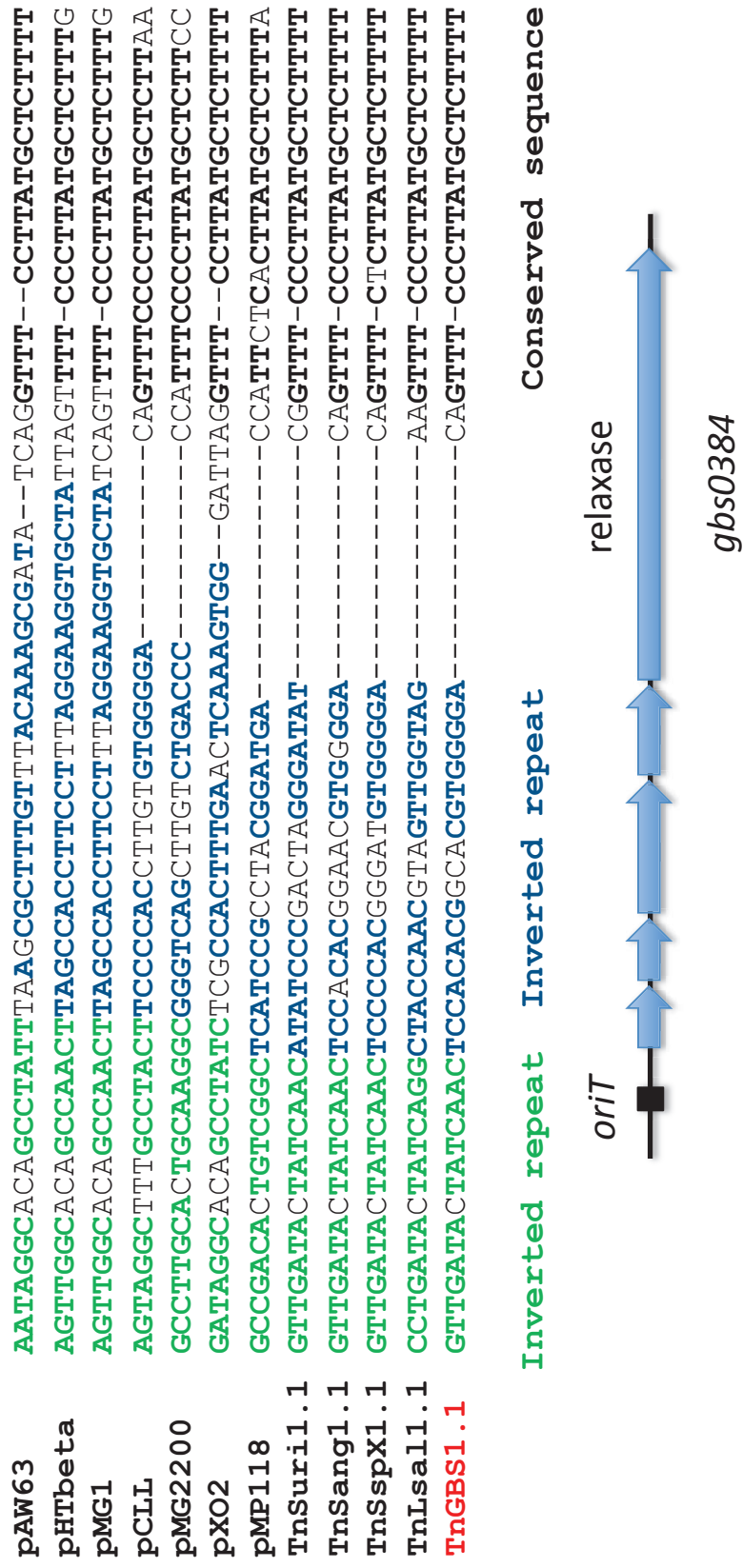


Figure 7



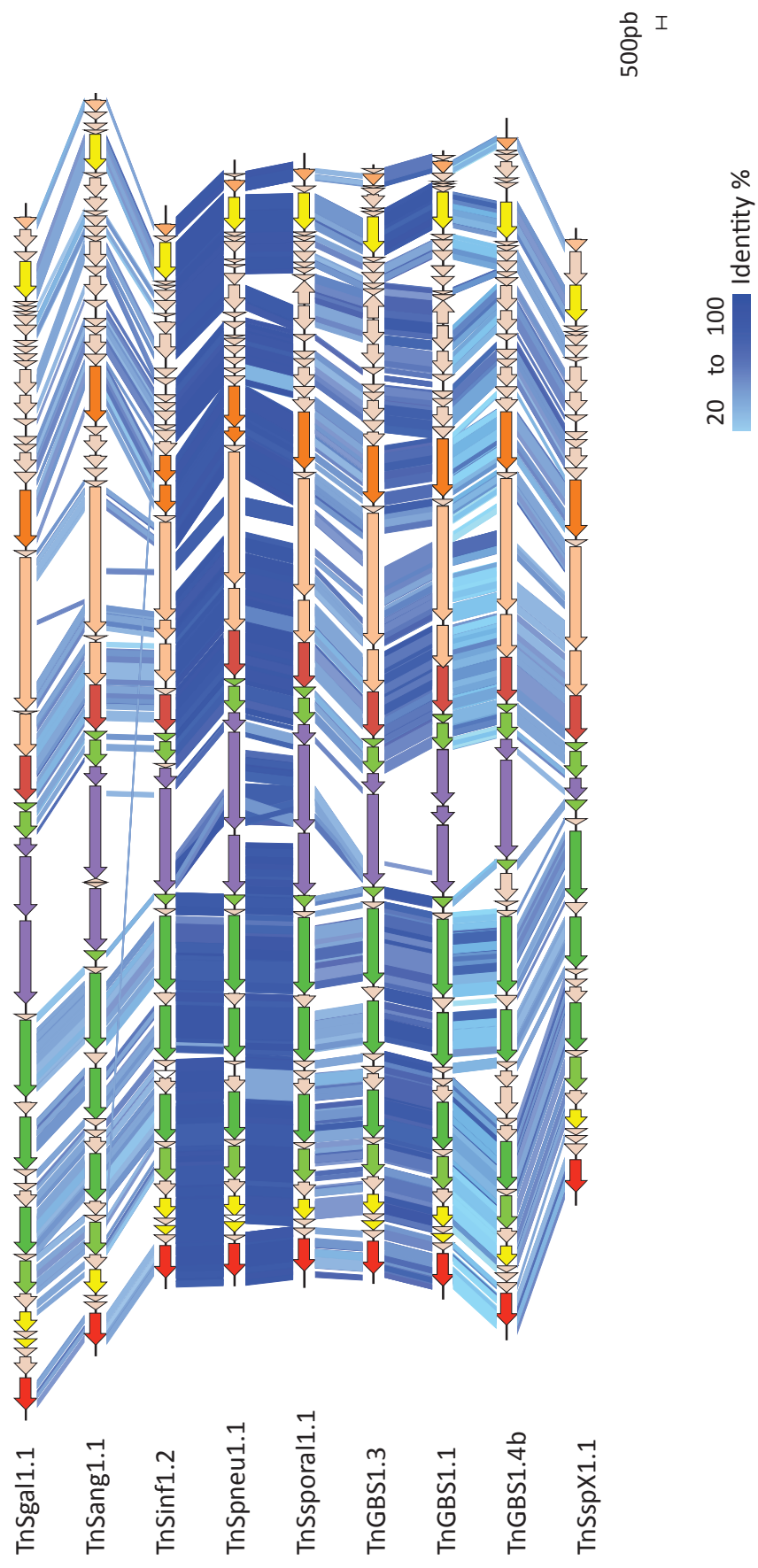


Figure S1

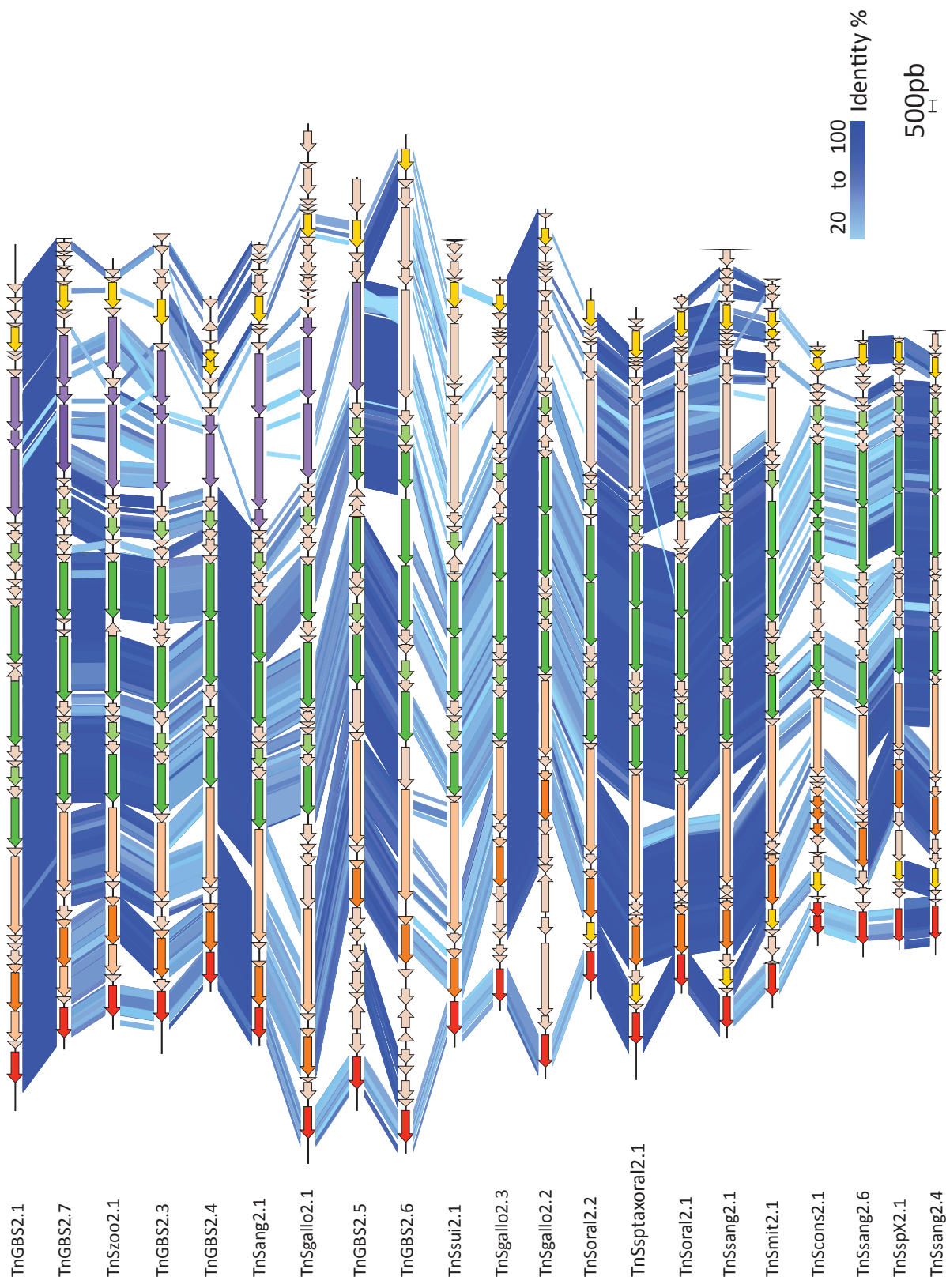


Figure S2



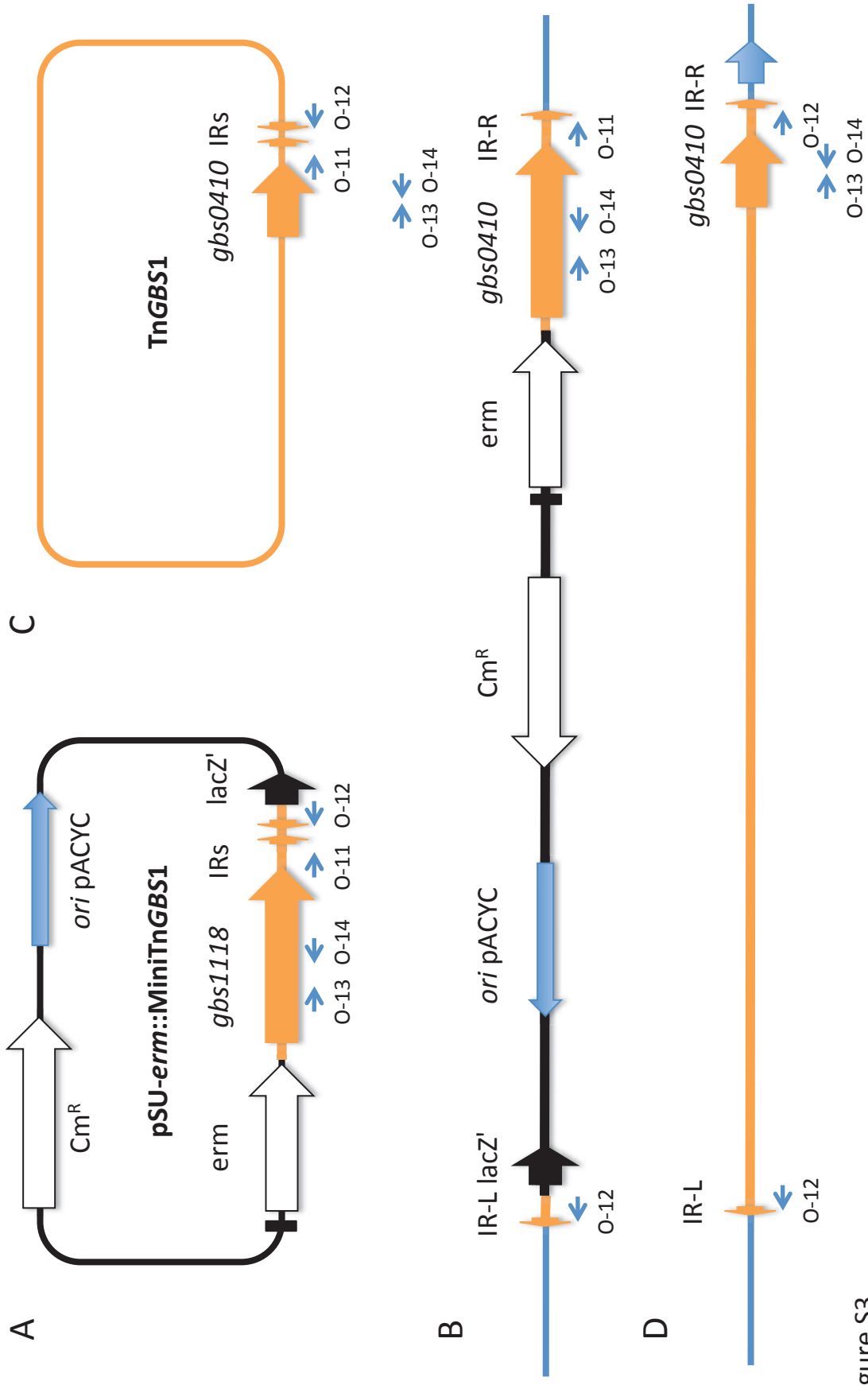


Figure S3

Targeted promoter	Insertion site	pSU::MiniTnGBS1			Total
		NEM316	BM110	TC BM110	
<i>gbs0057 - rpsJ</i>	ACATACAATAAAAAATAATGCTTGGCGTGTGCTGTTTTTTTTTGGATATCATATACTGG			2	2
<i>gbs0104 - pyrF</i>	ACATACTTTTTGCATAAAAACTTTGTAAACGTTTTAAATTAATGGTAAACTAACTGT			1	1
<i>gbs0125 - rpsL</i>	ACATACATTTTCAGTAAAAACTAGCATTTTTTATAGAATTGTCAGAAATTAACATG	1			1
<i>gbs0154 - tyrS</i>	ACATACTCAGATTTAATAATAACTTGCATATCCATAAGAGTTTTGATATAGTTGACATT		1		1
<i>gbs0156 ribD</i>	ACATACCAAGAGAATAGAACTTGATTTATTGGTAAAAATAAGTTATAATGTATGTAA	2	2	1	5
<i>gbs0258 - glyQ</i>	ACATACAGAATAAGATAAATAGTAGACATAAAAGTCATTTTTTGGTATACTGCTCAG	2	1		3
<i>gbs0301 gmk</i>	ACATACAAAATAATAGGAGACCTTGAATAAAAGGTTTCTTTTTTGGTAAACTAATGAGA	5			5
<i>gbs0353</i>	ACATACAGAAATCAGCAAACCTTGGAAAGACAGAGATATTTCTGGTATAATATCTCTA	1			1
<i>gbs0411</i>	ACATACAAAATAAAGAAGATGATTTGACTAAAGCCTATAAAACAAGTAAAAAAGAGGAA		1		1
<i>gbs0443 - rmpA</i>	ACATACAACATTTAGAAAAGAAATGGAGTCTTTTAGGTTATTTATGGTATAATAAAACCA	1	1		2
<i>gbs0494</i>	ACATACCGATGTAGAAAGTGACTTTACTTATTTTTTAAAAAAGTGGTATAGTTTATTTG	2			2
<i>gbs0553 - pyrD</i>	ACATACCATTATCATTATTTTCTTTCAGCTCTATTAATCAAGTGAATAAATAGTAAAAA		1		1
<i>gbs0615</i>	ACATACCGATATGATATCGCTTTCTTCTTATGTTTTTTGGTAAACTAATTAAG		1		1
<i>gbs0767 ribD</i>	ACATACAAACACATTATTTTTTATTTGACAAATAAAAAAGAGTTTAGTAAAAAATACGTCA	4	2		6
<i>gbs0772</i>	ACATACATAGACTTTTTCACCATTGGAAAATAAGGGTAGAAAAAGGATAAATTAACAATT		1		1
<i>gbs0809</i>	ACATACAGTTTATTGAAAGCGCTTGGACAAGATTAGTAAGAAATGATAAATTAAGACGT			1	1
<i>gbs0838</i>	ACATACGTTTAGAAATAAAACACATTTGCTTTTCCATTCTTTTTTTTATGCTATTACCAT		1		1
<i>gbs0842</i>	ACATACAGAAATTAATTATAGGTATTTTTAGTCTAATTATTATAGCGATTCTTGCTTGG		1		1
<i>gbs0910-gbs0911</i>	ACATACTTTTAAAGCATTATCTTTATCTTTTCATCTGTTATTTGGTAAAAATATCTTAAA	1			1
<i>gbs0968</i>	ACATACAAAATGCTATTTTAAATAGACATTTTTTATTTTTTGTCTGTATAATAGACAAA	2		1	3
<i>gbs1031</i>	ACATACATAATTTGAAAAAAGATAGTCAAACTCTTTAATTTTTTTGATAAATAGTCTCTA	1			1
<i>gbs1082 - pyrF</i>	ACATACATAATTTAGTATTTTATTTGACAAAATGATTGTAACCTGTAGAAATACCTACAA	1	1	1	3
<i>gbs1111</i>	ACATACGCCATTGAAAACAGTGTGTACTTTTTTCAACATTAATGTTACAATGATGCCCA		1		1
<i>gss1118-gbs1117</i>	ACATACCTCCTTTGCTTCAGTCTTGGACCTATTGCGGTTTTTATGCTACAATAGCTGTTT	1			1
<i>gbs1223</i>	ACATACATAATCAGTTTAGTTAGTTAGTTGACAATAATAATAATAATAATAATAATAATA	1			1
<i>gbs1232</i>	ACATACATTATTCATCTTATTTTGGCGATTTCAATAGTGATAATGGTATATTTAAAGTAA			1	1
<i>gbs1308 scpB</i>	ACATACTTTTGTCATGTTTTAACCTTGCATTAATCTTCTAAAGTAGTATAATATATATA	1			1
<i>gbs1314 int</i>	ACATACCAAATTTGAAGAAACTTAGTAATTTCTACAAATAAATTTGGTAAAAATGAGGAAA	2			2
<i>gbs1385 ribD</i>	ACATACAAGTCTTTGTTCTCACATAGAAATATCTTTTTAAAAATCTGATAAATAAGACTTT	1			1
<i>gbs1468 rpoB</i>	ACATACAAAGACTAAGATTACTTAGTCTTTTTTTCATGAAAAATGATAAATAAAGTCAT			1	1
<i>gbs1497 dnaG</i>	ACATACAAGCTTTTGAAAAAATTTACACAGAAAAAGAAAAATGATAAATAAATCAAAATGTA	1			1
<i>gbs1499 rpsU</i>	ACATACAAGATATCATAAAAAATATTTGACAATCTCTTTTTTATTTGGGTATCCTAGTAAGTG	12	1	2	15
<i>gbs1626 tmk</i>	ACATACAACAATCAAGTTGTTGGGTGCTTTTTTTTGATAAATTTGATAGAAATAGTAGTAA	1			1
<i>gbs1637 cspR</i>	ACATACAAACTCATGAATAGGTTGACTTTGTAGACTGAAGTGTATAATAGTGTCAAC		1		1
<i>gbs1638</i>	ACATACAATAACACAGATAAACTTGTGACAAGCCCTAAATTAGGTTATAATTAACCTTT	1			1
<i>gbs1650</i>	ACATACTTTTGTTTTACATATATTTGTAACACTCTTCTTTTTTTTGGTATAATGAGTCATG			1	1
<i>gbs1662</i>	ACATACCATTGTCAAAGTTTGTGATTTTATAGACTATTTAAGATATAATGACCTGTG	1			1
<i>gbs1681</i>	ACATACCTTCTTGTCAATAAGGTTTCAATCAGCTTGAAATAATGATAAATAAACAAGA		1		1
<i>gbs1691</i>	ACATACATTTCCATTTTACCAATTTGCTATCAAAGTGCATAATAATAGAGTTAAAAATTA	1			1
<i>gbs1724 aroD</i>	ACATACATAATTTACATACACTCTTTTTTAAATGAGTCAGTTTTTGGTAAAAATGAACFGAA	1			1
<i>gbs1739 efp</i>	ACATACCAAATCAATAAATATTTGAGAAAGCAGGTTATTTGATAAATAAGGTTGAT	1	1		2
<i>gbs1776</i>	ACATACATACATAAAGTCAAAATTTGACAGTTTCACTATTTTTTTGTTAGACTGGATATTA	1			1
<i>gbs1777 - glpF</i>	ACATACATTTTGTGATAAGCTTTTACAATATTAAGCTTACTGGTATAATGGTTAATG	1			1
<i>gbs1799</i>	ACATACATAATATCAAAAACGCTTGGTTTTGATATTTTTTATTTGGTATAATAAATGAAT	1			1
<i>gbs1814 - rpsL</i>	ACATACATAAAGTAATAAAGCTTGGACACATCATGTGTAATAATGATACTATGATAAACG	1	1		2
<i>gbs1835 - dltR</i>	ACATACATAATGCGGAAATTCGTAGACAATAGCCCTATTTTTTCTTAAAAATAGAGGA	1			1
<i>gbs1836 - rpmH</i>	ACATACAACCTGAAAGTATTTCTTGGACAGAGGCTAAAAATGTTATAAATAAGTATGG	1		1	2
<i>gbs1849</i>	ACATACCAAATAAAGAACACTTATTTTCAATTTTTCTTATTAATGATAAATAACGGATAA	1		1	2
<i>gbs1928 - relA</i>	ACATACCTTGGACCTCTAATCTTTTATTTTTAGGTTAAAAAGTATAAATAAGGTACA			1	1
<i>gbs2022</i>	ACATACCTTGAATCGGATTTATGTTTCAATCTTTTTTCTGTTTTGGTATAAATAGAAAA		1		1
<i>gbs2033</i>	ACATACAAATCAAAATTTAACTATTTGACAATTTATTTTTAAATGACGTAACCTTATAGGCA	1			1
<i>gbs2101 - gidA</i>	ACATACAGCCTATCCCTCTGGTTGACTTTTTAAAAAGATAAATGCTAAAAATAAATTTAAC	1			1
<i>rRNA 5S</i>	ACATACATATCAAAAAGAGATAATTTGACAACGTTACGGTTTCTTGGTATAGACTATAGATAT	2			2
<i>rRNA 16S</i>	ACATACACTTAAAAAAAAGTAGTTGACAAGCAGAATAAGTTGCTAGAAATATAGAAGT	1			1
<i>tRNA thr</i>	ACATACATTTCTCCAAATAGAATTTAGACTAAAAAAAACGGCTATTTTTTCTAAGTAAAGA	2			2
<i>tRNA cys</i>	ACATACATTTTACGAAAAGAGGCTTGGACGAATCTTTTTTTTTTCTGATAGAAATATTATCTG			1	1
		61	23	16	100

Figure S4

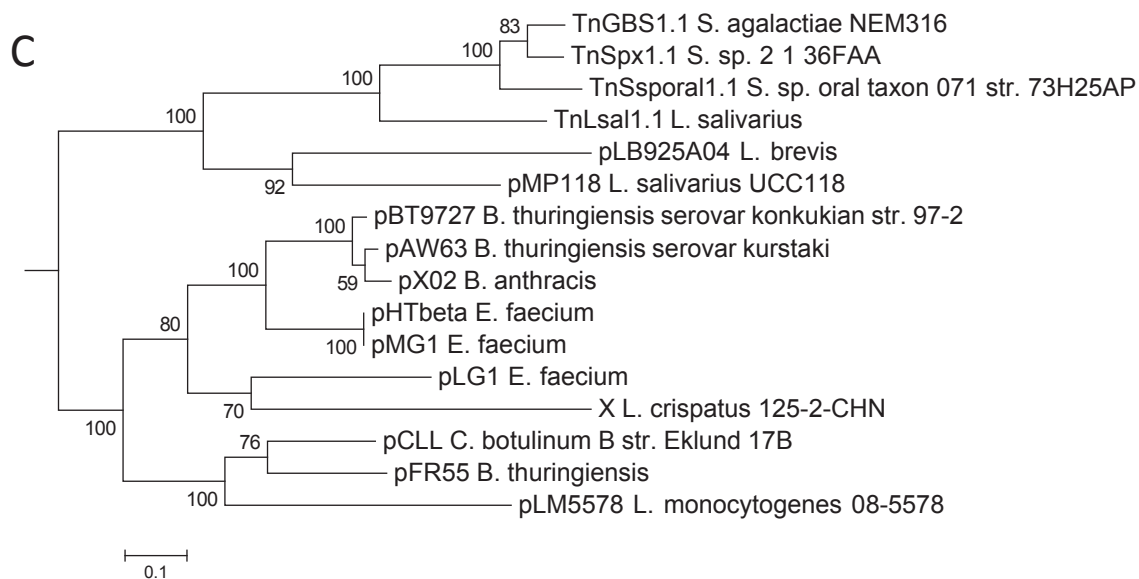
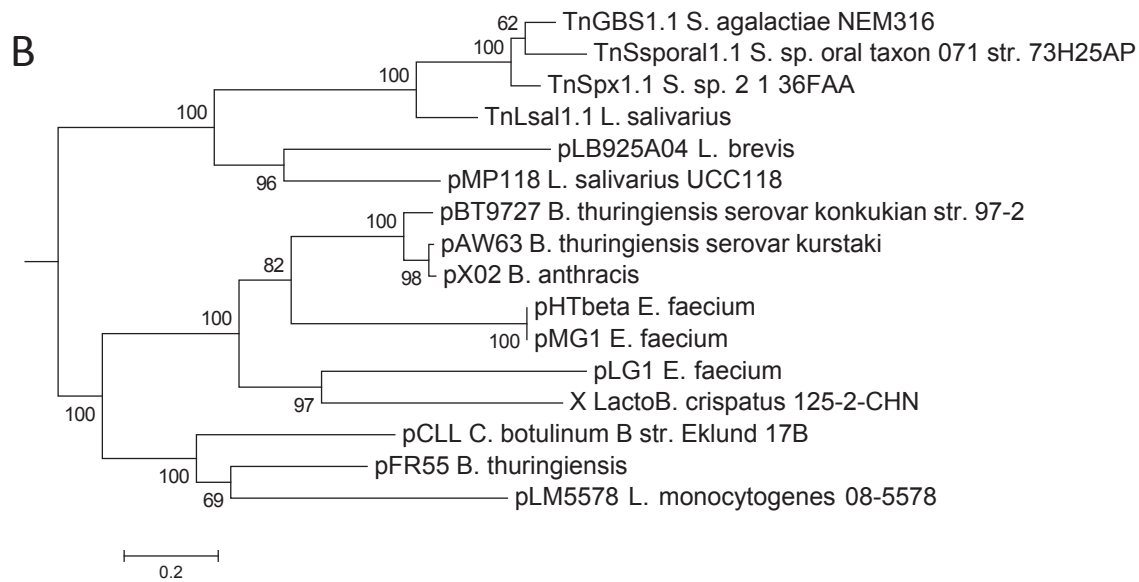
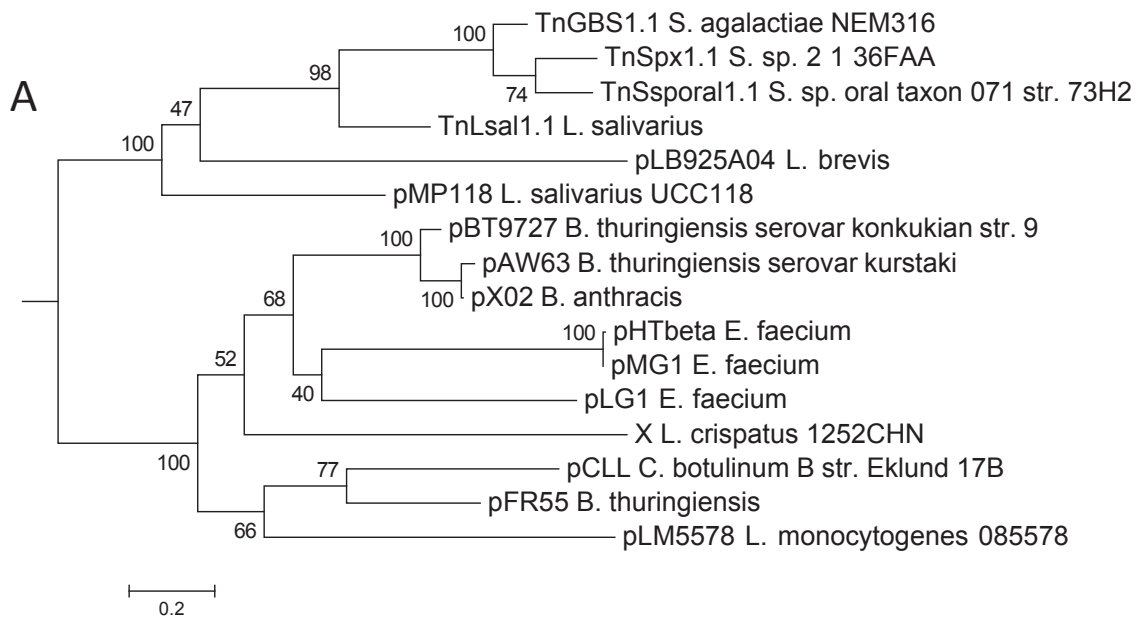


Figure S5

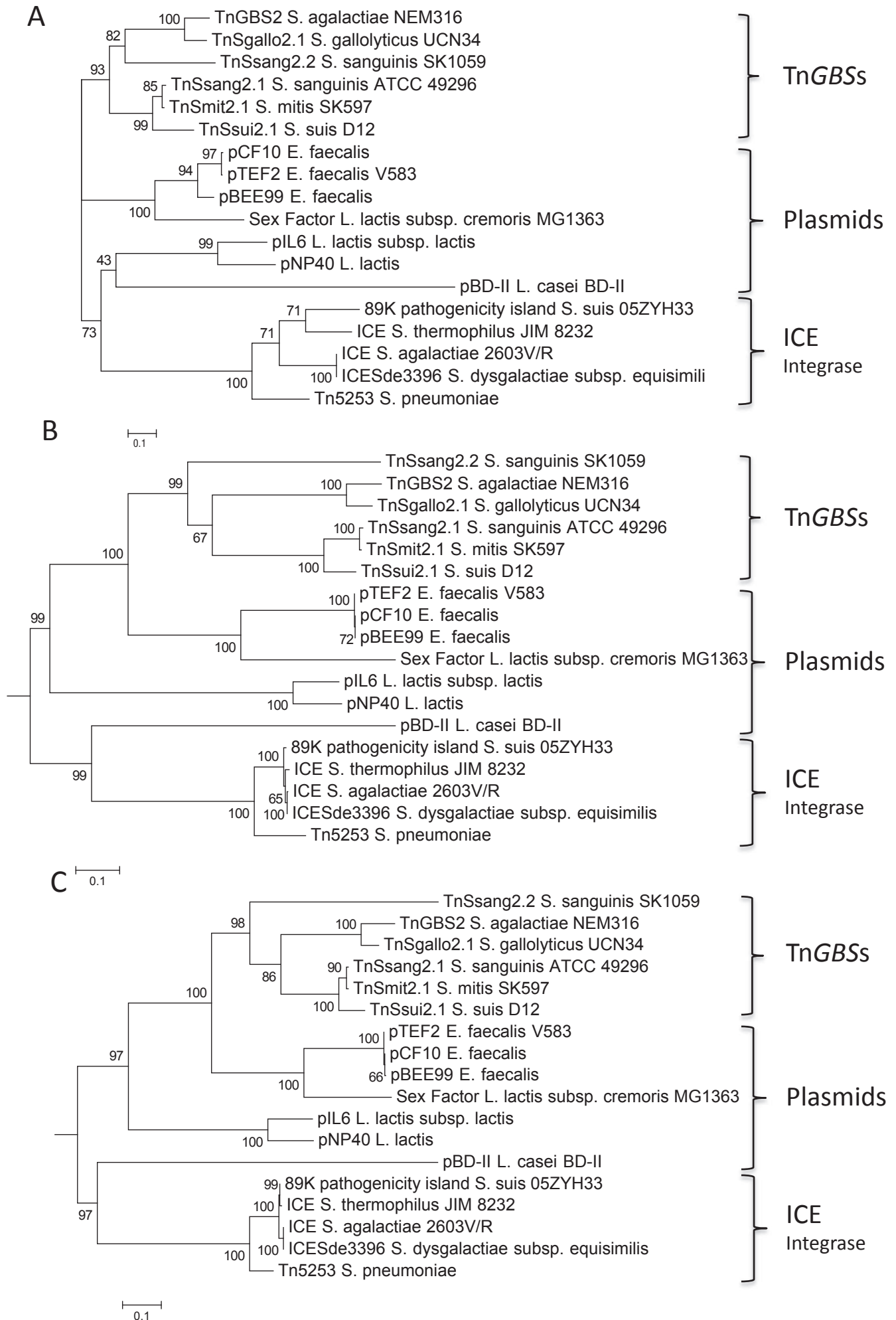
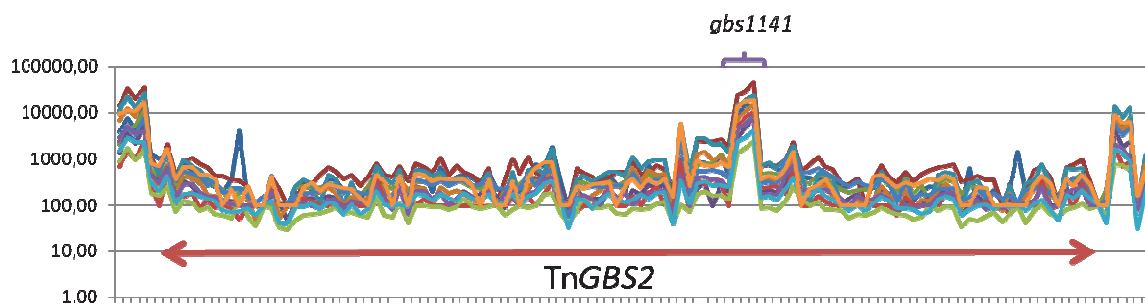


Figure S6

### 3. Données complémentaires de l'étude de la diversité des ICEs de la famille TnGBS.

Les données acquises jusqu'à présent, sur le processus de conjugaison des TnGBS, indiquent qu'au cours de ce processus, l'ICE passe par une succession d'états. Il est tout d'abord présent sous forme intégrée dans la bactérie donatrice, puis il s'excise et produit une forme circulaire. C'est cette forme circulaire qui va être reconnue par l'appareil de conjugaison et qui va être transférée dans la bactérie réceptrice. Une fois transféré à la bactérie réceptrice, l'ICE se re-circularise, se maintient pendant un moment par réplifications puis s'intègre. A chacun de ces états doit correspondre un profil d'expression particulier. Je me suis alors intéressée aux niveaux d'expressions des gènes de TnGBS2 dans la souche NEM316. L'analyse du niveau d'expressions de TnGBS2 m'a permis d'identifier des protéines et des petits ARNs potentiellement impliqués dans la régulation du processus de conjugaison des TnGBS.

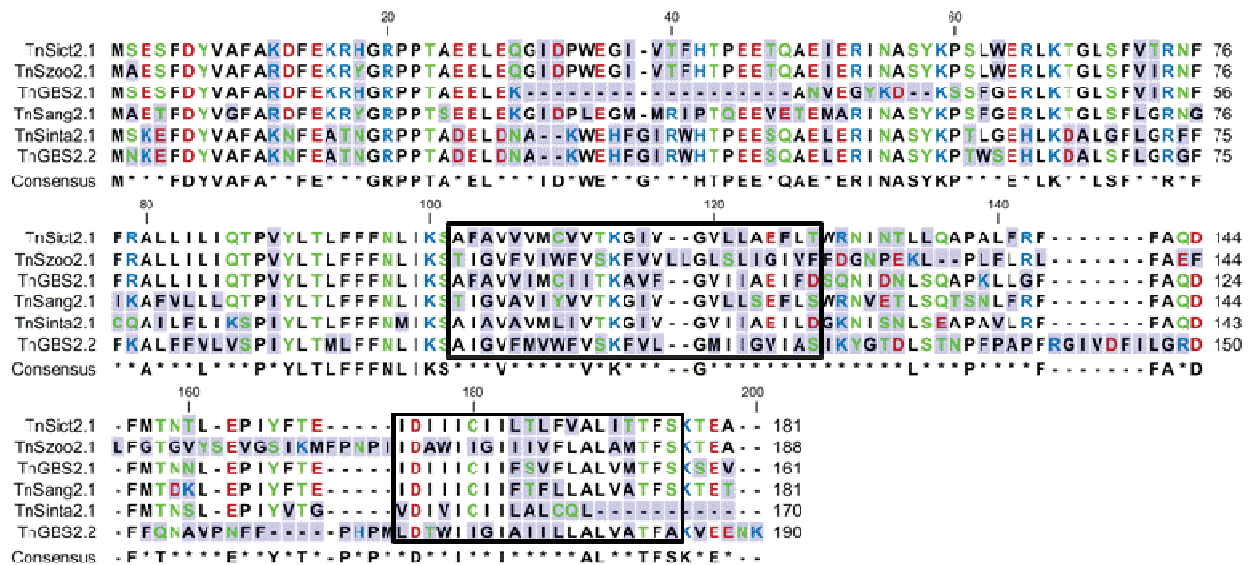
#### a) Expression des gènes de TnGBS1 et TnGBS2 dans la bactérie donatrice



**Figure 26: Expression des gènes de TnGBS2, analysée par microarrays, dans NEM316**  
Cette image représente en ordonnée les intensités de fluorescence observées pour les différentes sondes localisées au niveau de TnGBS2.

Pour avoir accès au niveau d'expression des gènes codés par TnGBS2, je me suis basée sur les intensités obtenues avec les sondes correspondantes à ces gènes, dans les expériences de microarrays. L'analyse, de 12 expériences transcriptomiques différentes réalisées dans la souche

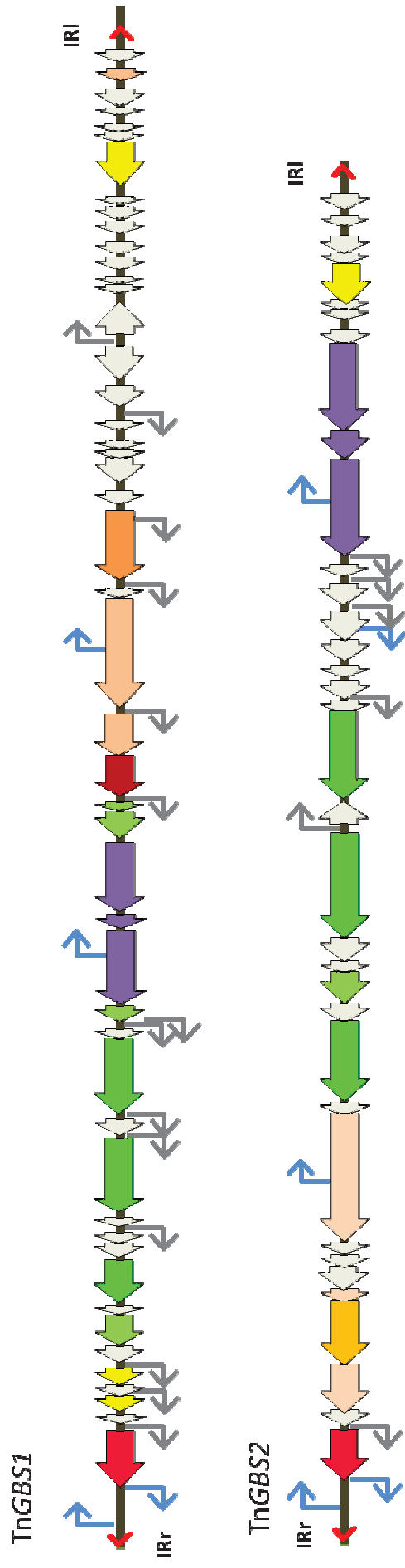
sauvage NEM316, révèle que les intensités observées pour les sondes correspondantes aux différents gènes de *TnGBS2* sont en général proches du bruit de fond, suggérant que la majorité des gènes de *TnGBS2* sont peu exprimés dans la souche NEM316 (Figure 26). Cette expression faible des gènes de *TnGBS2* permet probablement de limiter l'impact de *TnGBS2* sur la machinerie de transcription et de traduction de son hôte. Parmi les sondes dont l'intensité est supérieure au bruit de fond, on peut noter trois sondes correspondant au gène *gbs1141* et une sonde intergénique située entre *gbs1138* et *gbs1139*. Le gène *gbs1141* code pour une protéine de fonction inconnue, de 161 AA et présentant deux segments transmembranaires prédits par le programme TopPred [378, 379]. Cette protéine n'est pas conservée de manière uniforme au sein des éléments apparentés à *TnGBS2* (Figure 27). Néanmoins, son expression à un niveau plus élevé que le reste du transposon, et sa conservation au sein de certains ICEs apparentés, nous ont conduit à émettre l'hypothèse qu'elle pourrait être impliquée dans la régulation du processus de conjugaison, par exemple dans l'immunité de conjugaison suggérée par des résultats obtenus dans le paragraphe (II.B.4.c)



**Figure 27: Alignement des protéines apparentées à Gbs1141**

Les séquences protéiques correspondant aux deux domaines transmembranaires sont encadrées en noir. La conservation des acides aminés est indiquée par la coloration de la case, qui est le blanc pour les acides aminés conservés et bleu pour les acides aminés différents. L'écriture en rouge, bleu, vert et noir des acides aminés est dépendante de leur polarité.

Une expérience d'identification des sites d'initiation de la transcription a été réalisée par RNAseq au laboratoire par Isabelle Rosinski-Chupin. L'analyse des résultats a permis d'identifier 18 et 11 sites potentiels d'initiation de la transcription au sein respectivement, de Tn*GBS1* et de Tn*GBS2* (Figure 28), les sites d'initiation correspondant aux deux petits ARNs situés respectivement dans *gbs1126* et entre *gbs1138* et *gbs1139* (décrits dans le paragraphe suivant), ainsi que le site d'initiation du gène *gbs1141* présentant un nombre de lectures plus élevé que les autres promoteurs identifiés. Cette observation conforte l'idée que les deux petit ARNs et le gène *gbs1141* ont un niveau d'expression significatif lorsque Tn*GBS2* est intégré et *a priori* inactif.



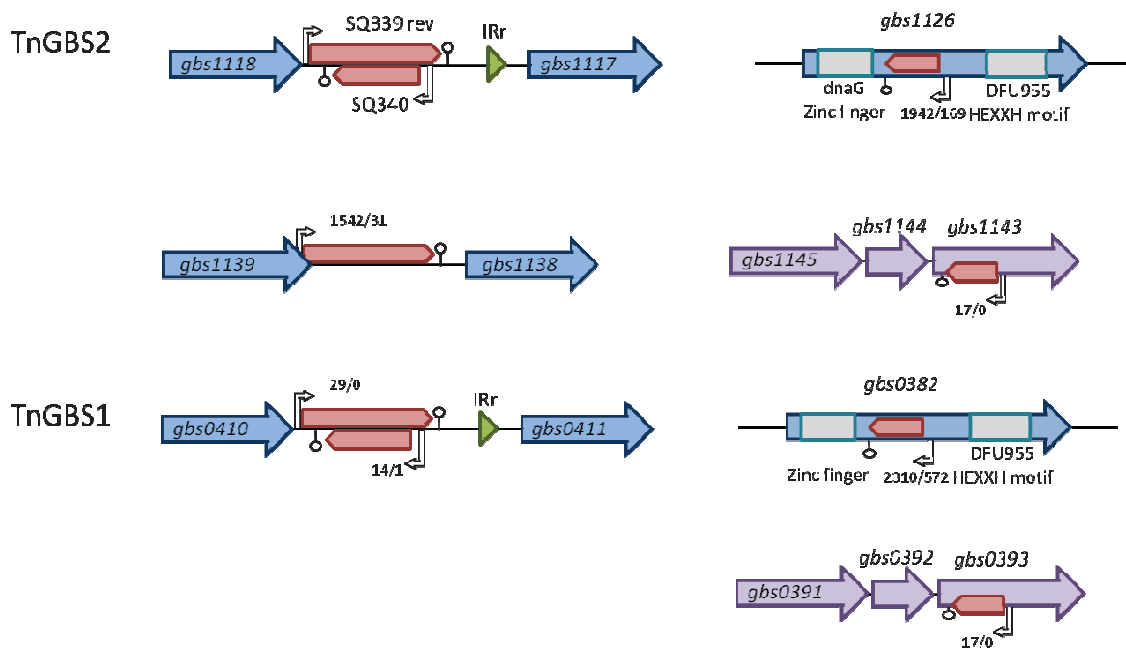
**Figure 28: Localisations des sites d'initiation de la transcription identifiés pour TnGBS1 et TnGBS2**

Les sites localisés en amont de gènes ou de petits ARNs non codants sont indiqués respectivement par des flèches grises et bleues. Les gènes codant pour la transposase sont indiqués en rouge. Les gènes codant pour des protéines impliquées dans la réplication sont en orange. Les gènes codant pour la relaxase, et des protéines impliquées dans la modification de l'ADN sont respectivement en orange foncé et orange clair. Le gène codant pour une protéine homologue à ClpA est en marron. Les protéines LPXTG sont codées par les gènes en violet. Les protéines contenant des domaines virB4, VirD4 et VirB6 sont codées par les gènes vert clair et d'autres protéines impliquées dans la formation du pore de conjugaison sont codées par des gènes colorés en vert foncé.



## b) Conservation de petits ARNs non codants entre Tn*GBS1* et Tn*GBS2*.

Les expériences d'identification de sites d'initiation de la transcription ont mises en évidence neuf sites d'initiation, qui pourraient être localisés en amont de petits ARNs non codants. Quatre de ces petits ARNs sont conservés entre Tn*GBS1* et Tn*GBS2*. Deux d'entre eux sont localisés entre l'IR droite et la transposase, un troisième est en anti-sens d'un des gènes codant pour une protéine LPXTG (*gbs1143*, *gbs0393*) et le quatrième est en anti-sens des gènes *gbs1126* et *gbs0386*. Le dernier ARN potentiel semble spécifique à Tn*GBS2* et est localisé entre les gènes *gbs1138* et *gbs1139* qui codent pour deux protéines de fonctions inconnues (Figure 29).



**Figure 29: Localisation des petits ARNs non codants potentiels de Tn*GBS1* et de Tn*GBS2***

La localisation et l'orientation des petits ARNs sont indiqués par des flèches rouges, les valeurs indiquées correspondent aux nombres de lecture obtenus lors de l'identification des sites d'initiation de la transcription tout d'abord avec un traitement par la TAP, puis en absence de traitement.

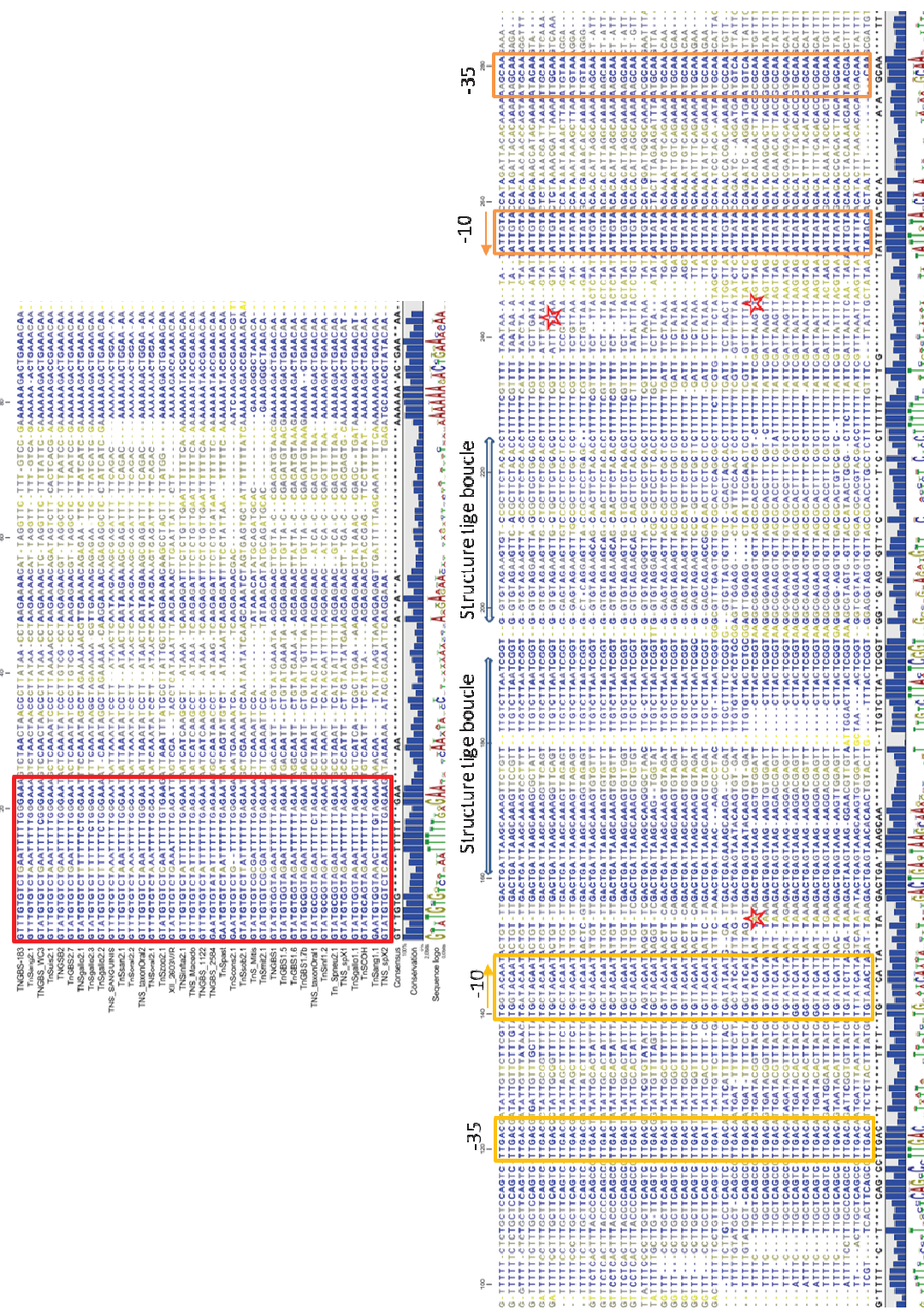
Les deux petits ARNs, situés entre l'IRr et la transposase de Tn*GBS2*, SQ339 et SQ340, sont localisés en anti-sens l'un par rapport à l'autre, et ont été aussi validés par Northern Blot par Christophe Pichon et al [380] à la suite de la prédiction *in silico* des petits

ARN de *S. agalactiae*. Les données d'identification des TSSs confirment la transcription de ces deux petits ARNs. Ces deux petits ARNs et leurs promoteurs sont conservés au sein des différents ICEs apparentés à *TnGBS1* et *TnGBS2* (Figure 30). Comme *TnGBS1* et *TnGBS2* correspondent à deux ICEs, qui codent pour des transposases apparentées, associés à des modules de conjugaison distincts, ce petit ARNs localisé entre l'IR et la transposase pourrait être impliqué dans le processus de transposition lui-même, ou dans sa régulation.

Un petit ARN, de 156 nucléotides, conservé entre *TnGBS1* et *TnGBS2*, est transcrit en anti-sens par rapport aux gènes respectifs *gbs0386* et *gbs1126*. Le gène *gbs0386* code pour une protéine avec un domaine DnaG retrouvé dans les protéines agissant comme primase. Le gène *gbs1126* code pour une protéine semblant fonctionner de manière modulaire, composée de 2 domaines. Dans certains des ICEs apparentés aux *TnGBS*, ou dans pCF10, ces deux domaines sont parfois codés par deux gènes distincts (article n°5). Des petits ARNs semblent également présents, en anti-sens des gènes orthologues, appartenant aux ICEs apparentés à *TnGBS1* et *TnGBS2*. Bien que les séquences de ces ARNs ne soient pas conservées, on peut identifier le promoteur et le terminateur potentiel de ces petits ARNs. La présence d'un petit ARN, en anti-sens d'un gène codant potentiellement pour une primase, suggère que ce petit ARN pourrait réguler l'activité de la primase ou son expression, ou qu'il pourrait agir comme matrice à la machinerie de réplication.

Le cinquième petit ARN, de 300nt, identifié dans *TnGBS2*, est transcrit au niveau de la région intergénique, localisée entre *gbs1139* et *gbs1138*, à partir d'un promoteur situé à la fin du gène *gbs1139*. L'existence d'une sonde correspondant à ce petit ARN sur nos microarrays permet de montrer que ce petit ARN a un niveau d'expression supérieur au bruit de fond dans la souche NEM316, confirmant sa transcription dans cette souche. Sa forte expression dans la souche NEM316 suggère que ce petit ARN pourrait être impliqué dans la régulation de l'expression de gènes de l'ICE.

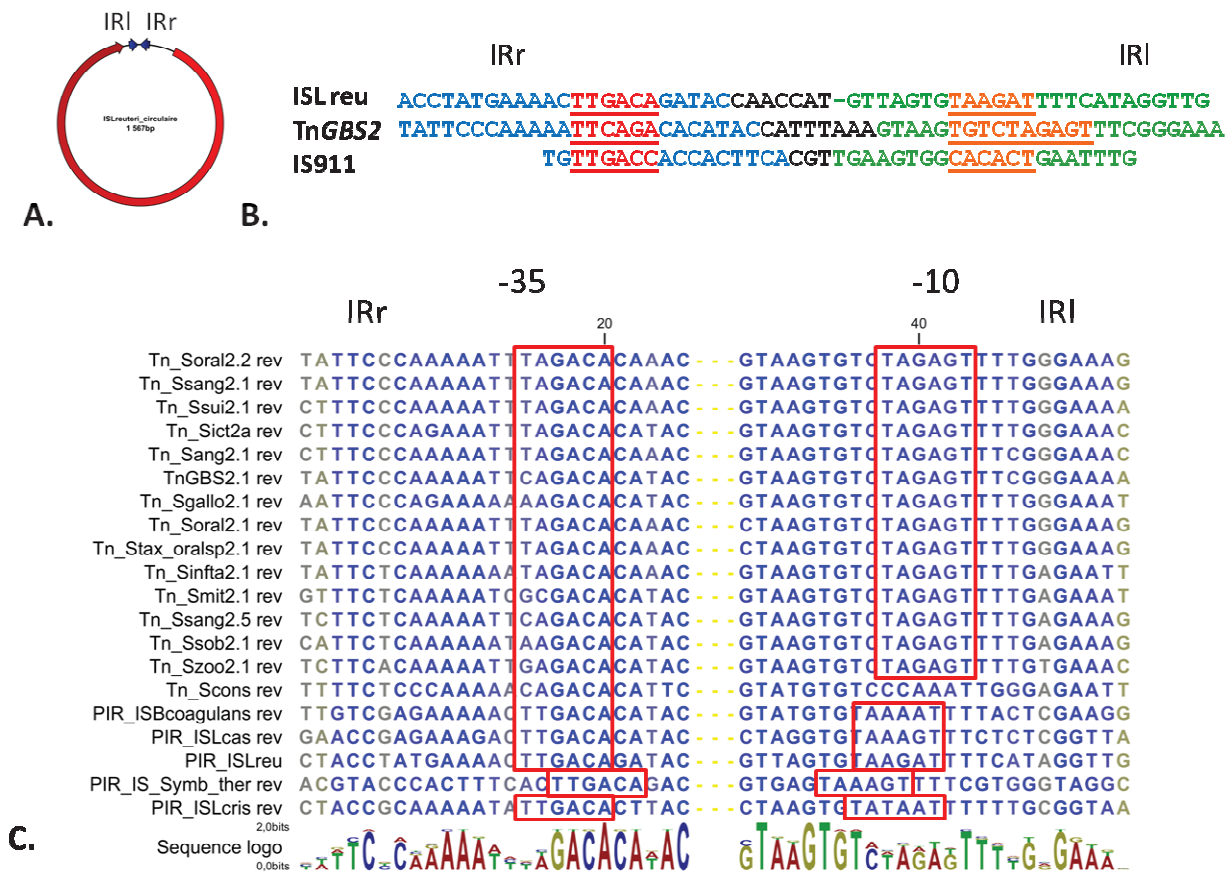
IRP



**Figure 30: Alignement de la région localisée entre l'IRP droite et la transposase et contenant deux petits ARNs non codants.**  
 La séquence correspondant à l'IRP est encadrée en rouge. Les séquences correspondant aux boîtes -35 et -10 des promoteurs de ces petits ARNs sont encadrées en orange, et les nucléotides localisés dans des étoiles rouges correspondent au +1 de transcription identifiés dans les expériences de RNAseq. Les nucléotides, impliqués dans la formation des deux structures tiges boucles prédites dans ces ARNs, sont indiqués par des flèches bleues.

**c) Les IRs de TnGBS2 forment un promoteur lors de la production de l'intermédiaire circulaire.**

La transposase des TnGBS, appartenant à une nouvelle famille de transposases à motif DDE, est également retrouvée dans des séquences d'insertion (IS) chez des Firmicutes (article n° 3). Dans le cas de l'IS911, appartenant à la famille IS3, il a été montré que la juxtaposition des 2 IRs dans la forme circulaire crée un promoteur,  $P_{\text{junc}}$ , contribuant à l'expression du gène codant pour la transposase [381]. De nombreuses autres séquences d'insertion, de la famille IS3, possèdent aussi un promoteur  $P_{\text{junc}}$  ou un équivalent [381]. L'analyse de la séquence des IRs des ISs apparentées à TnGBS chez *Lactococcus reuteri* montre que les ISs forment potentiellement un promoteur lors de la production de l'intermédiaire circulaire, les boîtes -35 et -10 étant portées respectivement par l'IR droite (IR<sub>r</sub>) et l'IR gauche (IR<sub>l</sub>). Ce promoteur semble conservé au sein d'autres ISs apparentées aux TnGBS, notamment chez *Bacillus coagulans*, *Lactobacillus crispatus*, *Lactococcus casei*, *Symbiobacterium thermophilum* et *L. helveticus*, et pourrait être également présent dans les ICEs apparentés à TnGBS2 (Figure 32 A, B et C). Cependant, ce promoteur n'est pas retrouvé dans les IRs des ICEs apparentés à TnGBS1 (Figure 32 C). Alors que dans les ISs ce promoteur contrôlerait l'expression de la transposase, dans les ICEs il pourrait être impliqué dans le contrôle de l'expression des gènes distaux de l'élément, permettant de distinguer l'ICE sous forme circulaire de l'ICE sous forme intégrée.



**Figure 31: P<sub>IR</sub> un promoteur également conservé dans les IRs des ISS, et des ICEs apparentés à TnGBS2.**

A. Visualisation de l'orientation des IRs au sein de la forme circulaire des ISS. B. Alignement des IRs de l'IS de *Lactobacillus reuteri*, de TnGBS2 et de l'IS911. C. Alignement Clustal des séquences nucléotidiques des IRs d'ISS et d'ICEs apparentés à TnGBS2. Les séquences correspondant potentiellement aux boîtes -35 et -10 sont encadrées en rouge.

#### 4. Analyse du processus de conjugaison de TnGBS2 et sa régulation

La caractérisation de la régulation de l'excision et du transfert des ICEs permet de mieux comprendre la propagation de ces éléments au sein des populations bactériennes. Les données de la littérature montrent que l'excision peut être un phénomène stochastique [309]. Elle peut aussi être inductible par des stimuli extérieurs comme la présence de bactéries réceptrices potentielles, ou l'exposition à des facteurs environnementaux en particulier aux mutagènes. La propagation des ICEs dépend aussi de leurs capacités à être maintenus au sein de leur hôte sans affecter de façon trop importante sa "fitness".

L'analyse de la fréquence de conjugaison de *TnGBS2*, et du nombre de formes circulaires de l'ICE dans la bactérie donatrice (article n°3 et 4), suggère que la production de l'intermédiaire circulaire est une étape limitante du transfert de *TnGBS2*. Cette observation permet de proposer qu'une augmentation du taux d'excision de *TnGBS2* conduirait probablement à une augmentation de l'efficacité de transfert de *TnGBS2*. L'identification de facteurs (protéines, ou conditions expérimentales) augmentant le taux d'excision de *TnGBS2* au sein de la bactérie donatrice permettrait de modifier l'efficacité de conjugaison de *TnGBS2*. Dans cette partie j'ai recherché des stimuli pouvant affecter la production de formes circulaires dans la bactérie donatrice ainsi que les gènes impliqués dans la modulation de l'efficacité de transfert de *TnGBS2*.

**a) Le transfert de *TnGBS2* n'est pas induit par la présence de bactéries potentiellement réceptrices.**

Afin de déterminer l'influence de bactéries potentiellement réceptrices sur la production de formes circulaires, j'ai quantifié ces formes circulaires dans la bactérie donatrice seule ou en présence de bactéries réceptrices. Les résultats obtenus indiquent que la présence de la bactérie réceptrice n'a pas d'effet significatif sur la quantité de formes circulaires (Table 14). Les résultats obtenus montrent que la quantité de formes circulaires semble constante chez la bactérie donatrice au cours du processus de conjugaison, et qu'elle n'est pas modifiée par la présence de la bactérie réceptrice. Ce qui conduit à suggérer que l'excision de *TnGBS2* dans la bactérie donatrice est constante et n'est pas modifiée par la présence de bactéries réceptrices sur le filtre.

temps	NEM316	NEM316 + BM110
0	$8.2 \times 10^{-4}$	$8.3 \times 10^{-4}$
1h	$7.1 \times 10^{-4}$	$9.5 \times 10^{-4}$
2h	$6.1 \times 10^{-4}$	$8.6 \times 10^{-4}$
3h	$6.1 \times 10^{-4}$	$6.1 \times 10^{-4}$
4h	$16 \times 10^{-4}$	$5 \times 10^{-4}$
nuite	$3.9 \times 10^{-4}$	$5.5 \times 10^{-4}$

**Tableau 14: Analyse de la production de formes circulaires en présence ou absence de bactéries réceptrices.**

Le ratio du nombre de formes circulaires sur le nombre de copies du transposon a été déterminé chez la donatrice seule et chez la donatrice en présence de la réceptrice.

## b) Le transfert de TnGBS2 n'est pas induit par la réponse SOS

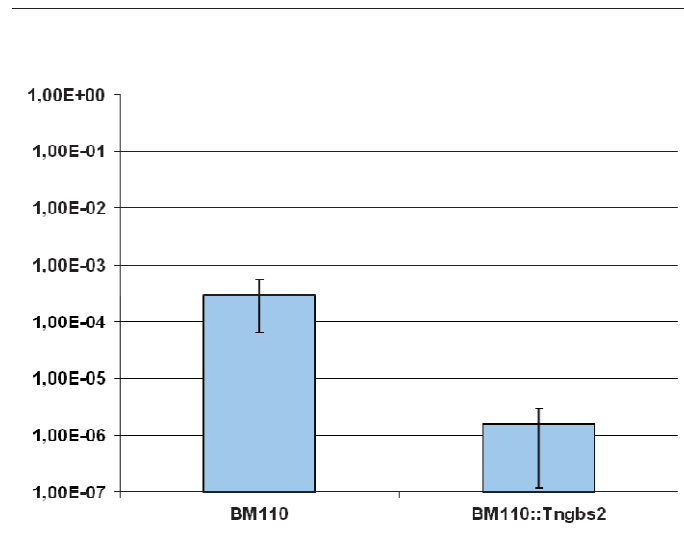
Pour certains éléments génétiques mobiles comme *ICES<sub>t1</sub>*, *ICES<sub>t3</sub>*, *ICEBs<sub>1</sub>*, et SXT, il a été montré qu'un traitement à la mitomycine C induit leur excision (*ICES<sub>t1</sub>*, *ICES<sub>t3</sub>*, *ICEBs<sub>1</sub>*) [304, 382], et/ou conduit à une augmentation de l'efficacité de leur transfert (*ICES<sub>t3</sub>*, de *ICEBs<sub>1</sub>* et SXT) [304, 312, 383]. Il a aussi été montré que le traitement à la mitomycine C induit l'expression de gènes requis pour le transfert de SXT [312]. La quantification par qPCR de la production de formes circulaires de TnGBS2, dans la souche NEM316, en réponse à l'exposition à différentes concentrations de mitomycine (Tableau 15), montre que la quantité de formes circulaires produites n'est pas affectée par des concentrations croissantes de mitomycine. Ces résultats suggèrent que l'excision et le transfert de TnGBS2 ne sont pas régulés par la réponse SOS de *S. agalactiae*.

**Tableau 15: Quantification de la production de formes circulaires de TnGBS2 en réponse à différentes concentrations de mitomycine C.**

		Concentration	0	20	50	100
		Mitomycine (ng/mL)				
Ratio quantité de forme circulaire/ quantité de transposon total	Phase Exponentielle		$1.3 \times 10^{-3}$ (0.5)	$1.4 \times 10^{-3}$ (0.3)	$1.4 \times 10^{-3}$ (0.2)	$1.5 \times 10^{-3}$ (0.2)
	Phase Stationnaire		$0.8 \times 10^{-3}$ (0.1)	$0.8 \times 10^{-3}$ (0.1)	$1.0 \times 10^{-3}$ (0.2)	$1.1 \times 10^{-3}$ (0.4)

Les valeurs indiquées correspondent au ratio de quantité de forme circulaire de l'ICE sur la quantité d'ICE total. les nombres indiqués entre des parenthèses correspondent à la déviation standard.

### c) Immunité de conjugaison et stabilité de Tn*GBS2* après conjugaison



**Figure 32: Immunité de conjugaison de Tn*GBS2***

Les fréquences de transfert indiquées, correspondent au transfert de Tn*GBS2* vers différentes souches réceptrices ne possédant pas Tn*GBS2*, comme BM110, ou possédant un élément identique à Tn*GBS2* comme BM110::Tn*GBS2*. Les valeurs indiquées correspondent à la moyenne de trois expériences indépendantes, et la déviation standard est donnée par une barre.

Le transfert de certains ICEs ou plasmides peut être affecté par la présence d'un élément identique, ou apparenté, dans la bactérie réceptrice, suggérant un processus d'immunité. Cette immunité peut intervenir au niveau de différentes étapes. (1) L'inhibition de la formation du pore entre les bactéries donatrice et réceptrice (exclusion de surface) [291, 292]. (2) L'inhibition de l'entrée de l'ADN dans la bactérie réceptrice (exclusion d'entrée) [293]. (3) L'inhibition de l'intégration (immunité de cible) [298-300]. (4) Une immunité liée à la présence d'un répresseur cI ou cI-like [301, 302].

Afin de déterminer si Tn*GBS2* pouvait être soumis à un tel phénomène d'immunité, j'ai testé l'efficacité de son transfert par conjugaison vers une souche réceptrice BM110 possédant, ou ne possédant pas une copie de Tn*GBS2* insérée. Les efficacités de transfert obtenues vers ces deux souches (Figure 32) montrent une diminution de la fréquence de transfert de Tn*GBS2* d'un facteur 100 lorsque la souche réceptrice possède Tn*GBS2*. Ce résultat met en évidence une immunité de conjugaison partielle. La diminution de la fréquence de conjugaison observée peut dépendre soit de la stabilité de l'appareil de transfert,



soit de la stabilisation de l'ICE dans le transconjugant. Le maintien de l'ICE dans le transconjugant est lié soit à son intégration dans le génome de la bactérie réceptrice, soit à sa réplication (article n°5).

## **5. Gbs1141, Gbs1147 et le promoteur P<sub>IR</sub>, trois éléments pouvant intervenir dans la régulation du processus de conjugaison**

Nous avons identifié deux éléments qui pourraient intervenir dans la régulation du processus de conjugaison. La protéine transmembraire Gbs1141 qui est exprimée à un niveau plus élevé que les autres gènes du transposon, et qui pourrait intervenir dans le phénomène d'immunité partielle décrit précédemment. Le promoteur, formé par les IRs lors de la production de l'intermédiaire circulaire, et qui, dans les ICEs pourrait être impliqué dans le contrôle de l'expression des gènes distaux de l'élément. À ces deux éléments s'ajoute la protéine Gbs1147 correspondant à un régulateur transcriptionnel de la famille cI et qui pourrait être impliquée dans le maintien à bas bruit de l'expression des gènes codant pour l'appareil de transfert, ou la transposase. Afin de déterminer un rôle potentiel de ces trois éléments dans la régulation du processus de conjugaison de Tn*GBS2*, j'ai construit des mutants de délétion des gènes *gbs1141* et *gbs1147* de Tn*GBS2* dans la souche NEM316, puis j'ai analysé l'effet de ces délétions sur la production de l'intermédiaire circulaire, et sur l'efficacité de transfert de Tn*GBS2*. J'ai également cherché à déterminer si le promoteur P<sub>IR</sub> créé par la juxtaposition des IRs dans la forme circulaire est fonctionnel et si son expression est régulée par Gbs1147.

Les expériences de quantification de la production de forme circulaire et la détermination de l'efficacité de transfert de Tn*GBS2* dans le mutant *gbs1141* indiquent que la protéine correspondante n'est impliquée, ni dans la régulation de la production de forme circulaire dans la bactérie donatrice, ni dans le processus de conjugaison lorsque la bactérie réceptrice ne possède pas d'éléments apparentés. Pour l'instant aucun phénotype ne peut donc être associé à ce mutant. Cependant, il reste à tester le rôle dans un contexte d'immunité de conjugaison.

**Tableau 16: Quantification de la production de formes circulaires de Tn*GBS2* dans le mutant *Δgbs1141*.**

Souche	Quantification forme circulaire
NEM316 phase exponentielle	$1.3 \times 10^{-3}$ (0.5)
NEM316 phase stationnaire	$0.8 \times 10^{-3}$ (0.1)
NEM316 <i>Δgbs1141</i> phase exponentielle	$2.4 \times 10^{-3}$ (0.5)
NEM316 <i>Δgbs1141</i> phase stationnaire	$2.2 \times 10^{-3}$ (0.7)

Les valeurs indiquées correspondent au ratio de quantité de forme circulaire de l'ICE sur la quantité d'ICE total. les nombres indiqués entre des parenthèses correspondent à la déviation standard.

Le gène *gbs1147* code pour une protéine présentant un domaine similaire au domaine de liaison à l'ADN du répresseur cI. Il a été montré que certaines protéines possédant ce domaine, comme SetR de SXT [313], ImmR de ICEBsl [303], et Arp1 de ICESt3 [382, 384], sont impliquées dans la répression transcriptionnelle de gènes portés par les éléments intégratifs correspondants, la mutation des répresseurs conduisant à une augmentation de l'excision de l'ICE [382]. La quantification de la production des formes circulaires de Tn*GBS2* dans le mutant *Δgbs1147* et dans le sauvage montre que, contrairement à ce qui a été observé pour l'ICESt3, la délétion du gène *gbs1147* de Tn*GBS2* n'affecte pas son excision (Tableau 17). Ces résultats sont en accord avec ceux obtenus dans des expériences de conjugaisons, qui montrent que la délétion de *gbs1147* n'affecte pas l'efficacité de transfert de Tn*GBS2* (données non montrées). Ce résultat est en accord avec ceux obtenus dans des expériences de conjugaison, qui montrent que la délétion de *gbs1147* n'affecte pas l'efficacité de transfert de Tn*GBS2* (données non montrées). Cette observation suggère que l'expression de la transposase n'est pas régulée par Gbs1147. Cependant ce régulateur pourrait avoir un rôle sur l'expression des autres gènes codés par Tn*GBS2* et sur l'expression du promoteur P<sub>IR</sub>.

**Tableau 17: Quantification de la production de formes circulaires de Tn*GBS2* dans le mutant *Δgbs1147***

Souche	Quantification forme circulaire
NEM316 phase exponentielle	$1.3 \times 10^{-3}$ (0.5)
NEM316 phase stationnaire	$0.8 \times 10^{-3}$ (0.1)
NEM316 <i>Δgbs1147</i> phase exponentielle	$2.8 \times 10^{-3}$ (0.4)
NEM316 <i>Δgbs1147</i> phase stationnaire	$3.4 \times 10^{-3}$ (0.5)

Les valeurs indiquées correspondent au ratio de quantité de forme circulaire de l'ICE sur la quantité d'ICE total. les nombres indiqués entre des parenthèses correspondent à la déviation standard.

Afin de valider la présence d'un promoteur potentiel dans les IRs de Tn*GBS2*, un fragment PCR amplifié à partir de la forme circulaire de Tn*GBS2* a été cloné en amont du gène *lacZ* dans le vecteur pTCVlac. Un dosage de l'activité bêta-galactosidase a été réalisé pour cette construction et pour les deux constructions témoins, qui correspondent au promoteur potentiel en orientation inverse et au vecteur vide. Ces dosages ont été faits dans la souche NEM316 et dans la souche BM110. Alors qu'aucune activité n'est détectée au bout de 24h pour les constructions témoins correspondant au vecteur vide et au vecteur avec le promoteur en orientation inverse, une activité est observée pour le promoteur P<sub>IR</sub> au bout de respectivement 5 heures et 18 heures dans les souches BM110 et NEM316 (Tableau 19). Ces résultats indiquent que les IRs dans la forme circulaire permettent la création d'un promoteur P<sub>IR</sub> de faible activité. La différence d'activité observée entre les souches BM110 et NEM316 suggère qu'un élément codé dans NEM316 est impliqué dans la répression de l'expression de ce promoteur. Afin de tester si la protéine Gbs1147 pourrait être impliquée dans la répression du promoteur P<sub>IR</sub>, le dosage de l'activité bêta-galactosidase a été réalisé dans la souche mutante *Δgbs1147*. Les résultats obtenus montrent que l'expression de ce promoteur est augmentée dans le mutant *Δgbs1147*, indiquant que Gbs1147 régulerait directement ou non l'activité du promoteur P<sub>IR</sub>.

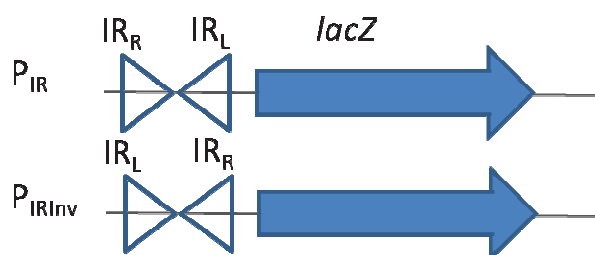


Figure 33: Orientation des promoteurs formés par les IRs dans les constructions  $P_{IR}$  et  $P_{IRInv}$

	pTCVlac vide	<sup>1</sup> $P_{IR}\Omega$ pTCVlac	$P_{IRInv}\Omega$ pTCVlac	<sup>2</sup> $P_{kana}\Omega$ pTCVlac
NEM316 stationnaire	ND	0.16 (0.12)	ND	1.67 (0.78)
BM110 stationnaire	ND	2.59 (0.78)	ND	NT
NEM316 $\Delta gbs1147$ stationnaire	ND	0.86 (0.18)	ND	NT

**Tableau 18: Activité béta-galactosidase du promoteur  $P_{IR}$  de *TnGBS2*.**

<sup>1</sup>L'activité est donnée en unités de Miller et la déviation standard est indiquée entre parenthèses. Les valeurs obtenues ont été calculées à partir de trois réplicats biologiques. <sup>2</sup>Le promoteur  $P_{kana}$  est utilisé comme promoteur contrôle. <sup>3</sup>ND signifie qu'aucune coloration jaune même très faible n'a été observée pour ces constructions au bout de 24h et <sup>4</sup>NT signifie que la construction n'a pas été testée.

Ces résultats permettent de proposer le modèle suivant dans lequel les IRs, lors de la production de l'intermédiaire circulaire, forment un promoteur transitoire qui contrôlerait l'expression des gènes distaux de *TnGBS2*. Dans la donatrice ce promoteur serait réprimé mais deviendrait actif après transfert. Parmi les gènes distaux sont retrouvés les gènes *gbs1149* et *gbs1147*, codant respectivement pour les protéines possédant un domaine RepA et un domaine cI. Le promoteur  $P_{IR}$  pourrait permettre l'expression transitoire de ces gènes juste après le transfert de *TnGBS2*, l'expression du régulateur cI conduisant par la suite à un rétro-contrôle négatif de l'activité du promoteur. L'expression transitoire de *gbs1149* pourrait être à l'origine du faible maintien de *TnGBS2* sous forme circulaire.

L'ensemble des résultats obtenus sur l'analyse du processus de conjugaison des *TnGBS*, et sa régulation a permis de montrer que la production de l'intermédiaire circulaire dans la donatrice est une étape limitante dans le transfert de *TnGBS2*. Les résultats obtenus montrent également que la production de la forme circulaire n'est pas induite par la présence de bactéries potentiellement réceptrices, et n'est pas modifiée par des perturbations modifiant

l'intégrité de l'ADN, comme la mitomycine. Ils montrent enfin que la seconde étape clé du processus de conjugaison est le maintien de l'ICE chez les transconjugants. Vis-à-vis de cette seconde étape, *TnGBS1* et *TnGBS2* semblent posséder des comportements très différents. Alors que *TnGBS2* est rapidement intégré dans le génome de la bactérie réceptrice, *TnGBS1* se maintient pendant de nombreuses générations sous forme circulaire. Ces deux profils de réplication versus intégration doivent refléter les deux programmes de régulation et d'expression distincts portés par *TnGBS1* et *TnGBS2*. Alors que *TnGBS2* possède un promoteur  $P_{IR}$  qui se forme lors de la production de la forme circulaire, et qui permet la distinction entre l'ICE sous forme intégrée et l'ICE sous forme circulaire, *TnGBS1* et l'ensemble des ICEs apparentés ne possèdent pas ce promoteur. Cette observation soulève la question de la régulation de l'expression du gène codant pour la protéine avec un domaine RepE potentiellement impliquée dans la réplication de *TnGBS1*. Il reste également à déterminer si d'autres éléments, tels que les petits ARNnc pourraient intervenir dans ce programme de régulation.

Une particularité des *TnGBS* est leur spécificité d'insertion originale au niveau des régions promotrices. Le maintien de ces ICEs au sein de la bactérie hôte pourrait s'expliquer par le fait que l'insertion des *TnGBS* au niveau des régions promotrices semble ne pas affecter l'expression des gènes adjacents, dans les conditions du laboratoire (article 3). Cependant, l'ensemble des sites d'insertion n'a pas été testé dans cette étude, et il n'est pas exclu que l'insertion à certains sites n'ait pas été sélectionnée positivement, ou du moins n'ait pas été contre-sélectionnée tout en modifiant l'expression du gène adjacent. En effet, certains facteurs de transcription agissent bien au delà de la région -35, et leur séquence de fixation peut ainsi se trouver très éloignée du promoteur par la présence de l'ICE. Par ailleurs, la présence d'une boîte -35, en bonne orientation dans la séquence de l'IR-droite des ICEs apparentés à *TnGBS2* et de certaines ISs, pourrait dans certains cas modifier l'expression du gène en aval en recréant un promoteur. Une telle modification d'expression du gène adjacent a été observée pour une IS apparentée dans *Lactobacillus helveticus*, ISLhe15, dont l'insertion en amont du promoteur du gène codant pour la bêta-galactosidase conduit à une expression constitutive de ce gène en absence de lactose [385].

## DISCUSSION ET PERSPECTIVES

Jusqu'à très récemment, l'analyse des réseaux de régulation était limitée à quelques organismes modèles comme *E. coli*, et *B. subtilis*. Avec le développement de nouvelles techniques d'études de l'expression génique comme les puces à ADN et le RNAseq, et leur coût de revient plus faible, les analyses transcriptomiques ont pu être étendues à un plus grand nombre d'espèces. Au cours de ma thèse, j'ai caractérisé, au moins partiellement, certains réseaux de régulation impliqués dans la modulation de l'expression de composants de la paroi de *S. agalactiae* grâce à des approches transcriptomiques. Les analyses transcriptomiques décrites précédemment dans *S. agalactiae* par J. MUSSER mettaient en évidence que des modifications de l'environnement comme la température, la phase de croissance ou la mise en culture dans des fluides biologiques, tels que le sang ou le liquide amniotique, conduisaient à la modification de l'expression d'un très grand nombre de gènes, pouvant aller jusqu'à plus de mille gènes, rendant difficile l'interprétation des résultats. Ces modifications d'expression résultent de l'action de nombreux régulateurs transcriptomiques qui peuvent être soit directement liés à la modification de l'environnement ou à la perturbation, mais elles peuvent également correspondre à des conséquences indirectes. Afin de caractériser comment *S. agalactiae* réagit à une absence de D-alanylation au niveau de la paroi, j'ai combiné une analyse globale de la réponse à la perturbation, et une approche génétique basée sur la mutation des régulateurs potentiellement impliqués dans cette réponse. Ceci m'a permis d'identifier des cibles directes des régulateurs transcriptionnels et de mettre en évidence certains effets probablement indirects. Cette analyse m'a également permis de montrer que l'expression des gènes, codant pour les protéines impliquées dans le processus de D-alanylation, étaient contrôlés de manière complexe et intégrés par plusieurs régulateurs, permettant probablement un ajustement fin des niveaux d'expression en fonction des conditions environnementales.

Ce type d'approche par analyse comparative des données transcriptomiques m'a également permis de mettre en évidence d'autres interconnexions entre les différents régulateurs de *S. agalactiae*, comme dans le cas des analyses de CiaRH avec RelRS et des régulateurs de la famille RofA. Cette dernière analyse a permis de démontrer une régulation croisée entre des régulateurs appartenant aux deux îlots génomiques SecA2/Srr et PI-2a, ainsi

que le contrôle indirect de l'expression des gènes de ces îlots par un régulateur appartenant au core génome CovRS.

L'acquisition de gènes, par transfert horizontal, permet à la bactérie d'acquérir rapidement de nouveaux traits phénotypiques comme la virulence, la résistance aux antibiotiques, et la capacité à utiliser de nouvelles sources d'énergie. Cependant l'ADN nouvellement acquis ne confère à la bactérie réceptrice un bénéfice que si les gènes apportés ne sont exprimés que de façon adéquate et coordonnée avec les autres gènes de la bactérie. Il a été constaté que lorsque l'ADN acquis contient un régulateur transcriptionnel, l'expression de ce régulateur est souvent contrôlée par un régulateur ancestral, comme cela a été observé pour l'îlot de pathogénicité de *Salmonella enterica* (SPI-2) [386], qui est régulée par un système à deux composants SsrB/SpiR situé dans l'îlot [387]. L'expression de ce TCS est contrôlée par les systèmes ancestraux OmpR/EnvZ et PhoP/Q présents dans le core génome [387].

Notre étude a également permis de montrer que certains régulateurs peuvent avoir des cibles légèrement différentes au sein des différentes souches de *S. agalactiae*. De même, une variabilité des réseaux de régulation parfois accompagnée de la pseudogénération de certains régulateurs, a été mise en évidence entre les souches. Ces observations peuvent expliquer la variabilité de présentation de structures de surface de *S. agalactiae*. Cette variabilité de réseaux de régulation est également potentiellement impliquée dans la diversité d'adaptation de *S. agalactiae*, suggérant que toutes les souches de *S. agalactiae* ne vont pas répondre de la même façon à des modifications d'environnement.

Enfin, l'analyse combinée de 26 transcriptomes a permis de prédire des cibles directes pour un certain nombre de régulateurs transcriptionnels chez *S. agalactiae*. La comparaison des cibles identifiées chez *S. agalactiae* aux cibles de régulateurs orthologues, chez d'autres espèces du genre streptocoque, indique que les régulateurs orthologues contrôlent l'expression de gènes impliqués dans des fonctions cellulaires communes, et donc que les réseaux de régulation semblent conservés au sein des espèces du genre streptocoque. Cependant on constate que cette conservation n'est pas totale. En effet, cette conservation n'est pas toujours vraie (i) au niveau du genre (ii) et que cette remarque peut être partiellement fautive également au niveau de l'espèce. Ces observations suggèrent que ces systèmes de régulation adoptés par ces différentes espèces sont adaptés à leur environnement. Ces modifications peuvent résulter des capacités de ces organismes à occuper de nouvelles niches écologiques, à contribuer à la

diversité phénotypique au sein des différentes espèces [388]. Cependant les modifications des systèmes de régulation ne résultent pas nécessairement du processus d'adaptation, même s'ils conduisent à des changements importants dans l'expression génique [389]. Vers la fin de ma thèse, j'ai eu accès aux données d'identification des sites d'initiation de la transcription. L'analyse préliminaire m'a permis de confirmer la conservation des sites de fixation de certains régulateurs (famille GntR et GlnR) au niveau des séquences promotrices des gènes ciblés. Grâce à ces données d'identification des sites d'initiation de la transcription, je vais pouvoir étendre mon analyse aux promoteurs des gènes regroupés en cluster, et rechercher la présence de motifs, conservés au niveau de ces promoteurs. Puis dans un second temps, j'étudierai la conservation de ces sites de fixation au sein des souches de *S. agalactiae* provenant de différents hôtes, mais également au sein du genre streptocoque. Il a été proposé que le gain ou la perte de sites de fixation (en opposition à des changements de séquences du facteur de transcription) est le mécanisme majeur contribuant à l'évolution des réseaux de régulation chez les eucaryotes supérieurs [390, 391]. Cette explication a été favorisée car elle permet d'altérer l'expression d'un seul gène sans affecter l'expression des gènes co-régulés par le même facteur de transcription. Cette vision n'est pas si simple dans le cas des bactéries, car il est convenu de l'importance des transferts horizontaux dans l'évolution des génomes bactériens.

Une étude préalable réalisée au sein du laboratoire suggérait que les transposons conjugatifs Tn*GBS1* et Tn*GBS2* sont impliqués dans des phénomènes de transferts de longs fragments d'ADN chromosomique par conjugaison, et que ces transferts pourraient avoir joués un rôle important dans l'évolution du génome des souches de *S. agalactiae* [236], en altérant l'aptitude de *S. agalactiae* à se multiplier au sein de son hôte et à le coloniser. Une des contraintes de la persistance des bactéries résultant de ces échanges dans la population pourrait être la compatibilité entre les différents réseaux de régulation provenant de la bactérie donatrice et de la bactérie réceptrice. L'organisation du génome bactérien en "module de transcription" avec co-localisation du régulateur avec les gènes cibles permet de limiter partiellement l'impact des échanges. Cependant, les résultats obtenus au cours de cette thèse, et les données de la littérature montrent que ce n'est pas toujours si simple. Les régulateurs et les gènes cibles peuvent être dispersés sur le chromosome. L'analyse des réseaux de régulation permettra de mieux appréhender l'impact de ces transferts de grands fragments d'ADN sur la bactérie.



La seconde partie de mon projet de thèse a portée sur l'étude de cette nouvelle famille d'ICE ou transposons conjugatifs, les Tn*GBS*. Cette famille constituait la première famille d'ICE dont la mobilité dépendait d'une transposase à motif DDE et non d'une intégrase. La seconde particularité de ces ICEs Tn*GBS* est leur spécificité d'insertion originale au niveau des régions promotrices. Nous avons montré que les ICEs de la famille Tn*GBS* peuvent être transférés au sein de l'espèce *S. agalactiae* où ils ont été identifiés. Ils semblent s'être également largement répandus au sein du genre streptocoque mais un seul membre a été identifié en dehors de ce genre, chez *L. salivarius*. Néanmoins, notre évaluation de la capacité de ces éléments à se répandre dans une espèce donnée pourrait encore être très biaisée par la nature des souches séquencées. Par exemple, les Tn*GBS* semblent être très répandus au sein des espèces *S. sobrinus* et *S. sanguinis*, mais semblent totalement absents de l'espèce *S. pyogenes*. Avec l'augmentation du nombre d'espèces bactériennes séquencées et du nombre de souches au sein de ces espèces, il est probable que de nouvelles séquences d'éléments apparentés aux Tn*GBS* seront retrouvées. Tn*GBS*1 et Tn*GBS*2 semblent avoir joué un rôle important dans l'évolution de *S. agalactiae*, en favorisant l'échange de grands fragments d'ADN chromosomique. L'identification d'un ICE de la famille Tn*GBS* dans le génome d'une souche de *S. pneumoniae* et de plusieurs souches de *S. mitis* et *S. oralis* soulève la question d'un rôle possible de ces ICEs dans les transferts génétiques, au sein de ces espèces qui représentent un problème important de santé publique.

La famille des Tn*GBS* est en réalité constituée de deux sous-familles, la sous-famille Tn*GBS*1 et la sous-famille Tn*GBS*2. Les membres de ces deux sous-familles possèdent la même transposase mais sont associés à des systèmes de conjugaison différents. Ce type de combinaison entre une transposase à motif DDE et un appareil de transfert n'a été jusqu'ici observé que pour deux familles d'ICEs, les Tn*GBS* et l'ICE6013 de *S. aureus* [392]. Cependant, d'autres familles d'ICEs, dont la mobilité dépendrait d'une transposase à motif DDE, existent probablement. Afin de vérifier cette hypothèse, il serait intéressant d'utiliser le même type d'approche, en recherchant des modules de conjugaison à proximité de transposase (à motif DDE) et en déterminant les limites des éléments grâce aux IRs, en commençant tout d'abord par les transposases de la famille de l'IS3 (contenant l'IS911) qui sont également capables de produire des intermédiaires de transposition circulaires [381]. Ce type d'approche permettra probablement d'identifier de nouvelles familles d'ICEs.

L'analyse comparative des conservations protéiques de *TnGBS1* et de *TnGBS2* a permis de mieux définir les protéines impliquées dans les différentes phases du processus de conjugaison: transposition, mobilisation, transfert et réplication. Elle a également montré que les membres de la sous-famille *TnGBS1* présentent une plus grande conservation protéique, suggérant que la sous-famille *TnGBS1* serait apparue plus récemment que la sous-famille *TnGBS2*, après acquisition d'un nouveau module de conjugaison. Les ICEs des deux sous-familles diffèrent aussi par leurs modules de régulation. Ceci explique probablement pourquoi *TnGBS1* et *TnGBS2* présentent certains comportements très différents : alors que *TnGBS2* est rapidement intégré dans le génome de la bactérie réceptrice, *TnGBS1* se maintient pendant de nombreuses générations sous forme circulaire. *TnGBS2* et *TnGBS1* pourraient aussi différer dans leurs capacités à induire un phénomène d'immunité de conjugaison: *TnGBS1* est retrouvé en copies multiples identiques ou non dans plusieurs souches de *S. agalactiae* comme NEM316 et FSL S3-026. Au contraire *TnGBS2* n'a jusqu'ici été observé que sous forme de copie unique. Cette observation est compatible avec le fait que nous ayons pu montrer expérimentalement une immunité partielle de conjugaison pour *TnGBS2*. Les analyses comparatives des différents membres de la famille des *TnGBS* ont aussi permis d'identifier des acteurs potentiels de la régulation de *TnGBS2* (comme le promoteur des IRs, les petits ARNs et les protéines Gbs1141 et Gbs1147). Les données préliminaires obtenues sur l'étude du processus de conjugaison de *TnGBS2* et de sa régulation m'ont conduites à proposer un modèle dans lequel l'expression génétique de *TnGBS2* dépendrait de l'état de l'ICE sous forme intégrée ou circulaire. Ce modèle serait cohérent mais différent de celui observé pour *Tn916* et les ICEs proches, montrant que la circularisation de l'ICE est requise pour l'expression du module de transfert. En effet, il a été montré que le promoteur contrôlant l'expression du module de conjugaison se situe de part et d'autre du site attI présent dans la forme circulaire de l'élément [242].

L'analyse comparative des processus de conjugaison de *TnGBS1* et de *TnGBS2* ainsi que de leurs programmes de régulation devrait permettre d'identifier des conditions stimulant leurs transferts et les mécanismes de cette stimulation. En retour, ces résultats permettront de mieux comprendre la propagation des *TnGBS* au sein du genre streptocoque. Ces éléments étant impliqués dans des phénomènes de transfert de longs fragments d'ADN chromosomique par conjugaison [236], il sera important de déterminer si les mêmes stimuli induisent à la fois le transfert des ICEs par conjugaison et les transferts de grand fragments d'ADN. Ceci

permettra de mieux comprendre les mécanismes senseurs contrôlant le taux de réarrangement chromosomique et le transfert horizontal de l'information génétique.

### 3. BIBLIOGRAPHIE

1. Nocard, E. and H. Mollereau, *Sur une mammite contagieuse des vaches laitières*. Annales de l'Institut Pasteur, 1887. **1**: p. 109-127.
2. Lehrmann K.B. and N. R., *Atlas und Grundriss der Bakteriologie und Lehrbuch der Speziellen bakteriologischen Diagnostik.*, ed. e.J.F. Lehmann. Vol. Teil Zwei - Text. 1896, München. 126.
3. Lancefield, R.C., *A serological differentiation of human and other groups of streptococci*. Journal of Experimental Medicine, 1933. **59**(141-158).
4. Lancefield, R.C. and E.H. Freimer, *Type-specific polysaccharide antigens of group B streptococci*. J Hyg (Lond), 1966. **64**(2): p. 191-203.
5. Glazunova, O.O., D. Raoult, and V. Roux, *Partial recN gene sequencing: a new tool for identification and phylogeny within the genus Streptococcus*. Int J Syst Evol Microbiol, 2010. **60**(Pt 9): p. 2140-8.
6. Tapp, J., M. Thollessen, and B. Herrmann, *Phylogenetic relationships and genotyping of the genus Streptococcus by sequence determination of the RNase P RNA gene, rnpB*. Int J Syst Evol Microbiol, 2003. **53**(Pt 6): p. 1861-71.
7. Eickhoff, T.C., et al., *Neonatal Sepsis and Other Infections Due to Group B Beta-Hemolytic Streptococci*. N Engl J Med, 1964. **271**: p. 1221-8.
8. McCracken, G.H., Jr., *Group B streptococci: the new challenge in neonatal infections*. J Pediatr, 1973. **82**(4): p. 703-6.
9. Baker, C.J., et al., *Suppurative meningitis due to streptococci of Lancefield group B: a study of 33 infants*. J Pediatr, 1973. **82**(4): p. 724-9.
10. Stoll, B.J., et al., *Early onset neonatal sepsis: the burden of group B Streptococcal and E. coli disease continues*. Pediatrics, 2011. **127**(5): p. 817-26.
11. Poyart, C., et al., *Invasive group B streptococcal infections in infants, France*. Emerg Infect Dis, 2008. **14**(10): p. 1647-9.
12. Krohn, M.A., S.L. Hillier, and C.J. Baker, *Maternal peripartum complications associated with vaginal group B streptococci colonization*. J Infect Dis, 1999. **179**(6): p. 1410-5.
13. Johri, A.K., et al., *Group B Streptococcus: global incidence and vaccine development*. Nat Rev Microbiol, 2006. **4**(12): p. 932-42.

14. Jourdan-Da Silva, N., D. Antona, C. Six, S. Georges, V. Goulet, P. Judlin, and D. Lévy-Bruhl., *Infections néonatales à streptococques B en France : données d'incidence de 1997 à 2006 et pratiques de prévention en maternité*. Bulletin Epidemiologique Hebdomadaire., 2008. **14-15** p. 110-113.
15. Tazi, A., et al., *The surface protein HvgA mediates group B streptococcus hypervirulence and meningeal tropism in neonates*. J Exp Med, 2010. **207**(11): p. 2313-22.
16. Weindling, A.M., et al., *Colonisation of babies and their families by group B streptococci*. Br Med J (Clin Res Ed), 1981. **283**(6305): p. 1503-5.
17. Hansen, S.M., et al., *Dynamics of Streptococcus agalactiae colonization in women during and after pregnancy and in their infants*. J Clin Microbiol, 2004. **42**(1): p. 83-9.
18. Lowell, A. and M.D. Rantz, *Suppurative arthritis due to hemolytic streptococcus of Lancefield group B: case report*. . Annals of Internal Medicine, 1940. **13**: p. 1744-1747.
19. Blancas, D., et al., *Group B streptococcal disease in nonpregnant adults: incidence, clinical characteristics, and outcome*. Eur J Clin Microbiol Infect Dis, 2004. **23**(3): p. 168-73.
20. Bergseng, H., et al., *Invasive group B streptococcus (GBS) disease in Norway 1996-2006*. Eur J Clin Microbiol Infect Dis, 2008. **27**(12): p. 1193-9.
21. Edwards, M.S. and C.J. Baker, *Group B streptococcal infections in elderly adults*. Clin Infect Dis, 2005. **41**(6): p. 839-47.
22. Tazi, A., et al., *Invasive group B streptococcal infections in adults, France (2007-2010)*. Clin Microbiol Infect, 2011.
23. Sendi, P., L. Johansson, and A. Norrby-Teglund, *Invasive group B Streptococcal disease in non-pregnant adults : a review with emphasis on skin and soft-tissue infections*. Infection, 2008. **36**(2): p. 100-11.
24. Farley, M.M., *Group B streptococcal disease in nonpregnant adults*. Clin. Infect. Dis., 2001. **33**(4): p. 556-561.
25. Phares, C.R., et al., *Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005*. Jama, 2008. **299**(17): p. 2056-65.
26. Edwards, M.S., et al., *Group B streptococcal colonization and serotype-specific immunity in healthy elderly persons*. Clin Infect Dis, 2005. **40**(3): p. 352-7.

27. Baker, C.J., *Group B streptococcal infections*. Clin Perinatol, 1997. **24**(1): p. 59-70.
28. Poyart, C., et al., *Centre National de Référence des Streptocoques Rapport d'activité 2009*, Institut de veille sanitaire
29. van der Mee-Marquet, N., et al., *Molecular characterization of human-colonizing Streptococcus agalactiae strains isolated from throat, skin, anal margin, and genital body sites*. J Clin Microbiol, 2008. **46**(9): p. 2906-11.
30. Yamamoto, T., et al., *Sexual transmission and reinfection of group B streptococci between spouses*. J Obstet Gynaecol Res, 1999. **25**(3): p. 215-9.
31. Daley, A.J. and S.M. Garland, *Prevention of neonatal group B streptococcal disease: progress, challenges and dilemmas*. J Paediatr Child Health, 2004. **40**(12): p. 664-8.
32. Yancey, M.K., et al., *The accuracy of late antenatal screening cultures in predicting genital group B streptococcal colonization at delivery*. Obstet Gynecol, 1996. **88**(5): p. 811-5.
33. Schuchat, A. and J.D. Wenger, *Epidemiology of group B streptococcal disease. Risk factors, prevention strategies, and vaccine development*. Epidemiol Rev, 1994. **16**(2): p. 374-402.
34. Lin, F.Y., et al., *Level of maternal IgG anti-group B streptococcus type III antibody correlated with protection of neonates against early-onset disease caused by this pathogen*. J Infect Dis, 2004. **190**(5): p. 928-34.
35. Baker, C.J., et al., *Immunization of pregnant women with a polysaccharide vaccine of group B streptococcus*. N Engl J Med, 1988. **319**(18): p. 1180-5.
36. Paoletti, L.C. and D.L. Kasper, *Glycoconjugate vaccines to prevent group B streptococcal infections*. Expert Opin Biol Ther, 2003. **3**(6): p. 975-84.
37. Rappuoli, R., *Reverse vaccinology*. Curr Opin Microbiol, 2000. **3**(5): p. 445-50.
38. Maione, D., et al., *Identification of a universal Group B streptococcus vaccine by multiple genome screen*. Science, 2005. **309**(5731): p. 148-150.
39. Brochet, M., et al., *Genomic diversity and evolution within the species Streptococcus agalactiae*. Microbes Infect, 2006. **8**(5): p. 1227-43.
40. Lindahl, G., M. Stalhammar-Carlemalm, and T. Areschoug, *Surface proteins of Streptococcus agalactiae and related proteins in other bacterial pathogens*. Clin. Microbiol. Rev., 2005. **18**(1): p. 102-127.
41. Keefe, G.P., *Streptococcus agalactiae mastitis: a review*. Can. Vet. J., 1997. **38**(7): p. 429-437.

42. Sorensen, U.B., et al., *Emergence and Global Dissemination of Host-Specific Streptococcus agalactiae Clones*. MBio, 2010. **1**(3).
43. Lebeuf Y, M.J.-C., Moineau S., *Ingrédients laitiers*. Science et technologie du lait, 2002.
44. Wilkinson, H.W., L.G. Thacker, and R.R. Facklam, *Nonhemolytic group B streptococci of human, bovine, and ichthyic origin*. Infect Immun, 1973. **7**(3): p. 496-8.
45. Eldar, A., et al., *Experimental streptococcal meningo-encephalitis in cultured fish*. Vet Microbiol, 1995. **43**(1): p. 33-40.
46. Kawamura, Y., et al., *High genetic similarity of Streptococcus agalactiae and Streptococcus difficilis: S. difficilis Eldar et al. 1995 is a later synonym of S. agalactiae Lehmann and Neumann 1896 (Approved Lists 1980)*. Int J Syst Evol Microbiol, 2005. **55**(Pt 2): p. 961-5.
47. Yildirim, A.O., C. Lammler, and R. Weiss, *Identification and characterization of Streptococcus agalactiae isolated from horses*. Vet Microbiol, 2002. **85**(1): p. 31-5.
48. Elliott, J.A., R.R. Facklam, and C.B. Richter, *Whole-cell protein patterns of nonhemolytic group B, type Ib, streptococci isolated from humans, mice, cattle, frogs, and fish*. J Clin Microbiol, 1990. **28**(3): p. 628-30.
49. Lammler, C., A. Abdulmawjood, and R. Weiss, *Properties of serological group B streptococci of dog, cat and monkey origin*. Zentralbl Veterinarmed B, 1998. **45**(9): p. 561-6.
50. Bruckler, J., I.W. Wibawan, and C. Lammler, *CAMP-reaction among skin isolates obtained from a dog with an acute squamous eczema*. Zentralbl Veterinarmed B, 1990. **37**(10): p. 767-9.
51. Schenkman, D.I., et al., *Outbreak of group B streptococcal meningoenzephalitis in athymic mice*. Lab Anim Sci, 1994. **44**(6): p. 639-41.
52. Slotved, H.C., et al., *Serotype IX, a Proposed New Streptococcus agalactiae Serotype*. J Clin Microbiol, 2007. **45**(9): p. 2929-36.
53. Bingen, E., et al., *Analysis of DNA restriction fragment length polymorphism extends the evidence for breast milk transmission in Streptococcus agalactiae late-onset neonatal infection*. J Infect Dis, 1992. **165**(3): p. 569-73.

54. Tamura, G.S., et al., *Analysis of restriction fragment length polymorphisms of the insertion sequence IS1381 in group B Streptococci*. J Infect Dis, 2000. **181**(1): p. 364-8.
55. Dmitriev, A., et al., *The presence of insertion elements IS861 and IS1548 in group B streptococci*. Folia Microbiol (Praha), 2003. **48**(1): p. 105-10.
56. Dmitriev, A., et al., *ISSa4-based differentiation of Streptococcus agalactiae strains and identification of multiple target sites for ISSa4 insertions*. J Bacteriol, 2004. **186**(4): p. 1106-9.
57. Jones, N., et al., *Multilocus sequence typing system for group B Streptococcus*. J. Clin. Microbiol., 2003. **41**(6): p. 2530-2536.
58. Diedrick, M.J., et al., *Clonal analysis of colonizing group B Streptococcus, serotype IV, an emerging pathogen in the United States*. J Clin Microbiol, 2010. **48**(9): p. 3100-4.
59. Manning, S.D., et al., *Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada*. J Clin Microbiol, 2009. **47**(4): p. 1143-8.
60. Eickel, V., et al., *Emergence of respiratory Streptococcus agalactiae isolates in cystic fibrosis patients*. PLoS One, 2009. **4**(2): p. e4650.
61. Gherardi, G., et al., *Molecular epidemiology and distribution of serotypes, surface proteins, and antibiotic resistance among group B streptococci in Italy*. J Clin Microbiol, 2007. **45**(9): p. 2909-16.
62. Jones, N., et al., *Enhanced invasiveness of bovine-derived neonatal sequence type 17 group B Streptococcus is independent of capsular serotype*. Clin. Infect. Dis., 2006. **42**(7): p. 915-924.
63. Luan, S.L., et al., *Multilocus sequence typing of Swedish invasive group B Streptococcus isolates indicates a neonatally associated genetic lineage and capsule switching*. J. Clin. Microbiol., 2005. **43**(8): p. 3727-33.
64. Martins, E.R., et al., *Analysis of group B streptococcal isolates from infants and pregnant women in Portugal revealing two lineages with enhanced invasiveness*. J. Clin. Microbiol., 2007. **45**(10): p. 3224-9.
65. Sadowy, E., B. Matynia, and W. Hryniewicz, *Population structure, virulence factors and resistance determinants of invasive, non-invasive and colonizing Streptococcus agalactiae in Poland*. J Antimicrob Chemother, 2010. **65**(9): p. 1907-14.



66. Marchaim, D., et al., *Clonal variability of group B Streptococcus among different groups of carriers in southern Israel*. Eur. J. Clin. Microbiol. Infect. Dis., 2006. **25**(7): p. 443-8.
67. Brochet, M., et al., *Population structure of human isolates of Streptococcus agalactiae from Dakar and Bangui*. J Clin Microbiol, 2009. **47**(3): p. 800-3.
68. Huber, C.A., et al., *Characterization of invasive and colonizing isolates of Streptococcus agalactiae in East African adults*. J Clin Microbiol, 2011. **49**(10): p. 3652-5.
69. Sun, Y., et al., *Comparison of a 3-set genotyping system with multilocus sequence typing for Streptococcus agalactiae (Group B Streptococcus)*. J Clin Microbiol, 2005. **43**(9): p. 4704-7.
70. Feil, E.J., et al., *eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data*. J Bacteriol, 2004. **186**(5): p. 1518-30.
71. Springman, A.C., et al., *Selection, recombination, and virulence gene diversity among group B streptococcal genotypes*. J Bacteriol, 2009. **191**(17): p. 5419-27.
72. Evans, J.J., et al., *Phylogenetic relationships among Streptococcus agalactiae isolated from piscine, dolphin, bovine and human sources: a dolphin and piscine lineage associated with a fish epidemic in Kuwait is also associated with human neonatal infections in Japan*. J Med Microbiol, 2008. **57**(Pt 11): p. 1369-76.
73. Bisharat, N., et al., *Hyperinvasive neonatal group B streptococcus has arisen from a bovine ancestor*. J. Clin. Microbiol., 2004. **42**(5): p. 2161-2167.
74. Hanage, W.P., C. Fraser, and B.G. Spratt, *The impact of homologous recombination on the generation of diversity in bacteria*. J Theor Biol, 2006. **239**(2): p. 210-9.
75. Brochet, M., *Génomique des populations et flux géniques au sein de l'espèce Streptococcus agalactiae*. 2007, Université Paris VI: Paris.
76. Héry-Arnaud, G., *Rôle des éléments génétiques mobiles dans l'évolution et la virulence de streptococcus agalactiae*. 2009, Université François-Rabelais de Tours.
77. Nizet, V. and C.E. Rubens, *Pathogenic mechanisms and virulence factors of group B streptococci*. Gram-positive pathogens, ed. V.A. Fischetti, et al. 2000, Washington, D.C.: ASM Press. 125-136.
78. Doran, K.S. and V. Nizet, *Molecular pathogenesis of neonatal group B streptococcal infection: no longer in its infancy*. Mol Microbiol, 2004. **54**(1): p. 23-31.

79. Tamura, G.S., et al., *Adherence of group B streptococci to cultured epithelial cells: roles of environmental factors and bacterial surface components*. *Infect Immun*, 1994. **62**(6): p. 2450-8.
80. Caillouette, J.C., et al., *Vaginal pH as a marker for bacterial pathogens and menopausal status*. *Am J Obstet Gynecol*, 1997. **176**(6): p. 1270-5; discussion 1275-7.
81. Pavletic, A.J., et al., *Experience with routine vaginal pH testing in a family practice setting*. *Infect Dis Obstet Gynecol*, 2004. **12**(2): p. 63-8.
82. Borges, S., J. Silva, and P. Teixeira, *Survival and biofilm formation by Group B streptococci in simulated vaginal fluid at different pHs*. *Antonie Van Leeuwenhoek*, 2011. **2011**: p. 29.
83. Butler, K.M., C.J. Baker, and M.S. Edwards, *Interaction of soluble fibronectin with group B streptococci*. *Infect Immun*, 1987. **55**(10): p. 2404-8.
84. Tamura, G.S., A. Nittayajarn, and D.L. Schoentag, *A glutamine transport gene, *glnQ*, is required for fibronectin adherence and virulence of group B streptococci*. *Infect Immun*, 2002. **70**(6): p. 2877-85.
85. Dramsi, S., et al., *Assembly and role of pili in group B streptococci*. *Mol Microbiol*, 2006. **60**(6): p. 1401-13.
86. Mistou, M.Y., et al., *Molecular dissection of the *secA2* locus of group B Streptococcus reveals that glycosylation of the *Srr1* LPXTG protein is required for full virulence*. *J Bacteriol*, 2009. **191**(13): p. 4195-206.
87. Sheen, T.R., et al., *Serine-Rich Repeat Proteins and Pili Promote Streptococcus agalactiae Colonization of the Vaginal Tract*. *J Bacteriol*, 2011. **193**(24): p. 6834-42.
88. Baron, M.J., et al., *Alpha C protein of group B Streptococcus binds host cell surface glycosaminoglycan and enters cells by an actin-dependent mechanism*. *J Biol Chem*, 2004. **279**(23): p. 24714-23.
89. Tyrrell, G.J., et al., *Binding and invasion of HeLa and MRC-5 cells by Streptococcus agalactiae*. *Microbiology*, 2002. **148**(Pt 12): p. 3921-31.
90. Pritchard, D.G., et al., *Characterization of the group B streptococcal hyaluronate lyase*. *Arch Biochem Biophys*, 1994. **315**(2): p. 431-7.
91. Christie, R., N.E. Atkins, and E. Munch-Petersen, *A note on a lytic phenomenon shown by group B streptococci*. *Australian Journal of Experimental Biology and Medical Science*, 1944. **22**: p. 197-200.

92. Nsagha, D.S., C.S. Bello, and Y.T. Kandakai-Olukemi, *Hippurate hydrolysis and Christie, Atkins, Munch-Peterson tests as epidemiological diagnostic tools for Streptococcus agalactiae carriage in pregnancy*. East Afr Med J, 2000. **77**(1): p. 34-6.
93. Skalka, B. and J. Smola, *Lethal effect of CAMP-factor and UBERIS-factor--a new finding about diffusible exosubstances of streptococcus agalactiae and Streptococcus uberis*. Zentralbl Bakteriol A, 1981. **249**(2): p. 190-4.
94. Lang, S. and M. Palmer, *Characterization of Streptococcus agalactiae CAMP factor as a pore-forming toxin*. J Biol Chem, 2003. **278**(40): p. 38167-73.
95. Pritzlaff, C.A., et al., *Genetic basis for the beta-haemolytic/cytolytic activity of group B Streptococcus*. Mol Microbiol, 2001. **39**(2): p. 236-247.
96. Cederlund, A., G.H. Gudmundsson, and B. Agerberth, *Antimicrobial peptides important in innate immunity*. Febs J, 2011. **278**(20): p. 3942-51.
97. Guani-Guerra, E., et al., *Antimicrobial peptides: general overview and clinical implications in human health and disease*. Clin Immunol, 2010. **135**(1): p. 1-11.
98. Zipfel, P.F. and C. Skerka, *Complement regulators and inhibitory proteins*. Nat Rev Immunol, 2009. **9**(10): p. 729-40.
99. Gros, P., F.J. Milder, and B.J. Janssen, *Complement driven by conformational changes*. Nat Rev Immunol, 2008. **8**(1): p. 48-58.
100. Poyart, C., et al., *Attenuated virulence of Streptococcus agalactiae deficient in D-alanyl-lipoteichoic acid is due to an increased susceptibility to defensins and phagocytic cells*. Mol Microbiol, 2003. **49**(6): p. 1615-25.
101. Hamilton, A., et al., *Penicillin-binding protein 1a promotes resistance of group B streptococcus to antimicrobial peptides*. Infect Immun, 2006. **74**(11): p. 6179-87.
102. Harris, T.O., et al., *A novel streptococcal surface protease promotes virulence, resistance to opsonophagocytosis, and cleavage of human fibrinogen*. J Clin Invest, 2003. **111**(1): p. 61-70.
103. Poyart, C., et al., *Contribution of Mn-cofactored superoxide dismutase (SodA) to the virulence of Streptococcus agalactiae*. Infect. Immun., 2001. **69**(8): p. 5098-5106.
104. Puopolo, K.M. and L.C. Madoff, *Upstream short sequence repeats regulate expression of the alpha C protein of group B Streptococcus*. Mol. Microbiol., 2003. **50**(3): p. 977-991.

105. Madoff, L.C., et al., *Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein*. Proc Natl Acad Sci U S A, 1996. **93**(9): p. 4131-6.
106. Cieslewicz, M.J., et al., *Structural and genetic diversity of group B streptococcus capsular polysaccharides*. Infect Immun, 2005. **73**(5): p. 3096-103.
107. Mavroidi, A., et al., *Genetic relatedness of the Streptococcus pneumoniae capsular biosynthetic loci*. J Bacteriol, 2007. **189**(21): p. 7841-55.
108. Cieslewicz, M.J., et al., *Functional analysis in type Ia group B Streptococcus of a cluster of genes involved in extracellular polysaccharide production by diverse species of streptococci*. J Biol Chem, 2001. **276**(1): p. 139-46.
109. Chaffin, D.O., L.M. Mentele, and C.E. Rubens, *Sialylation of group B streptococcal capsular polysaccharide is mediated by cpsK and is required for optimal capsule polymerization and expression*. J Bacteriol, 2005. **187**(13): p. 4615-26.
110. Hulse, M.L., et al., *Effect of type III group B streptococcal capsular polysaccharide on invasion of respiratory epithelial cells*. Infect Immun, 1993. **61**(11): p. 4835-41.
111. Gibson, R.L., et al., *Group B streptococci invade endothelial cells: type III capsular polysaccharide attenuates invasion*. Infect Immun, 1993. **61**(2): p. 478-85.
112. Wessels, M.R., et al., *Definition of a bacterial virulence factor: sialylation of the group B streptococcal capsule*. Proc Natl Acad Sci U S A, 1989. **86**(22): p. 8983-7.
113. Marques, M.B., et al., *Prevention of C3 deposition by capsular polysaccharide is a virulence mechanism of type III group B streptococci*. Infect Immun, 1992. **60**(10): p. 3986-93.
114. Paoletti, L.C., R.A. Ross, and K.D. Johnson, *Cell growth rate regulates expression of group B Streptococcus type III capsular polysaccharide*. Infect Immun, 1996. **64**(4): p. 1220-6.
115. Ross, R.A., L.C. Madoff, and L.C. Paoletti, *Regulation of cell component production by growth rate in the group B Streptococcus*. J Bacteriol, 1999. **181**(17): p. 5389-94.
116. Gray, B.M., and D. G. Pritchard., *Phase variation in the pathogenesis of group B streptococcal infections*. New perspectives on streptococci and streptococcal infections, ed. I.G. Orefici. 1992, Stuttgart, Germany: Gustav Fischer. p. 452–454.
117. Koskiniemi, S., M. Sellin, and M. Norgren, *Identification of two genes, cpsX and cpsY, with putative regulatory function on capsule expression in group B streptococci*. FEMS Immunol Med Microbiol, 1998. **21**(2): p. 159-68.

118. Greenberg, D.N., et al., *Sensitivity and specificity of rapid diagnostic tests for detection of group B streptococcal antigen in bacteremic neonates*. J Clin Microbiol, 1995. **33**(1): p. 193-8.
119. Davies, S., et al., *Streptococcus grouping latex kits: evaluation of five commercially available examples*. Br J Biomed Sci, 2003. **60**(3): p. 136-40.
120. Sutcliffe, I.C., G.W. Black, and D.J. Harrington, *Bioinformatic insights into the biosynthesis of the Group B carbohydrate in Streptococcus agalactiae*. Microbiology, 2008. **154**(Pt 5): p. 1354-63.
121. Pritchard, D.G., B.M. Gray, and H.C. Dillon, Jr., *Characterization of the group-specific polysaccharide of group B Streptococcus*. Arch Biochem Biophys, 1984. **235**(2): p. 385-92.
122. Michon, F., et al., *Structure of the complex group-specific polysaccharide of group B Streptococcus*. Biochemistry, 1987. **26**(2): p. 476-86.
123. Michon, F., et al., *Multiantennary group-specific polysaccharide of group B Streptococcus*. Biochemistry, 1988. **27**(14): p. 5341-51.
124. Michon, F., et al., *The alpha-L-(1----2)-trirhamnopyranoside epitope on the group-specific polysaccharide of group B streptococci*. Infect Immun, 1991. **59**(5): p. 1690-6.
125. Glaser, P., et al., *Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease*. Mol. Microbiol., 2002. **45**(6): p. 1499-1513.
126. Vallejo, J.G., C.J. Baker, and M.S. Edwards, *Roles of the bacterial cell wall and capsule in induction of tumor necrosis factor alpha by type III group B streptococci*. Infect Immun, 1996. **64**(12): p. 5042-6.
127. Lancefield, R.C., M. McCarty, and W.N. Everly, *Multiple mouse-protective antibodies directed against group B streptococci. Special reference to antibodies effective against protein antigens*. J Exp Med, 1975. **142**(1): p. 165-79.
128. Marques, M.B., et al., *Functional activity of antibodies to the group B polysaccharide of group B streptococci elicited by a polysaccharide-protein conjugate vaccine*. Infect Immun, 1994. **62**(5): p. 1593-9.
129. Weidenmaier, C. and A. Peschel, *Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions*. Nat Rev Microbiol, 2008. **6**(4): p. 276-87.

130. Fischer, W., T. Mannsfeld, and G. Hagen, *On the basic structure of poly(glycerophosphate) lipoteichoic acids*. *Biochem Cell Biol*, 1990. **68**(1): p. 33-43.
131. Toon, P., P.E. Brown, and J. Baddiley, *The lipid-teichoic acid complex in the cytoplasmic membrane of Streptococcus faecalis N.C.I.B. 8191*. *Biochem J*, 1972. **127**(2): p. 399-409.
132. Neuhaus, F.C. and J. Baddiley, *A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria*. *Microbiol Mol Biol Rev*, 2003. **67**(4): p. 686-723.
133. Doran, K.S., et al., *Blood-brain barrier invasion by group B Streptococcus depends upon proper cell-surface anchoring of lipoteichoic acid*. *J Clin Invest*, 2005. **115**(9): p. 2499-507.
134. Grundling, A. and O. Schneewind, *Synthesis of glycerol phosphate lipoteichoic acid in Staphylococcus aureus*. *Proc Natl Acad Sci U S A*, 2007. **104**(20): p. 8478-83.
135. Oku, Y., et al., *Pleiotropic roles of polyglycerolphosphate synthase of lipoteichoic acid in growth of Staphylococcus aureus cells*. *J Bacteriol*, 2009. **191**(1): p. 141-51.
136. Henneke, P., et al., *Role of lipoteichoic acid in the phagocyte response to group B streptococcus*. *J Immunol*, 2005. **174**(10): p. 6449-55.
137. Archibald, A.R., J. Baddiley, and S. Heptinstall, *The alanine ester content and magnesium binding capacity of walls of Staphylococcus aureus H grown at different pH values*. *Biochim Biophys Acta*, 1973. **291**(3): p. 629-34.
138. Fischer, W. and P. Rosel, *The alanine ester substitution of lipoteichoic acid (LTA) in Staphylococcus aureus*. *FEBS Lett*, 1980. **119**(2): p. 224-6.
139. Hurst, A., et al., *Loss of D-alanine during sublethal heating of Staphylococcus aureus S6 and magnesium binding during repair*. *J Gen Microbiol*, 1975. **89**(2): p. 277-84.
140. MacArthur, A.E. and A.R. Archibald, *Effect of culture pH on the D-alanine ester content of lipoteichoic acid in Staphylococcus aureus*. *J Bacteriol*, 1984. **160**(2): p. 792-3.
141. Poyart, C., et al., *Regulation of D-alanyl-lipoteichoic acid biosynthesis in Streptococcus agalactiae involves a novel two-component regulatory system*. *J Bacteriol*, 2001. **183**(21): p. 6324-34.
142. Fischer, W., P. Rosel, and H.U. Koch, *Effect of alanine ester substitution and other structural features of lipoteichoic acids on their inhibitory activity against autolysins of Staphylococcus aureus*. *J Bacteriol*, 1981. **146**(2): p. 467-75.

143. Steen, A., et al., *Autolysis of Lactococcus lactis is increased upon D-alanine depletion of peptidoglycan and lipoteichoic acids*. J Bacteriol, 2005. **187**(1): p. 114-24.
144. Hughes, A.H., I.C. Hancock, and J. Baddiley, *The function of teichoic acids in cation control in bacterial membranes*. Biochem J, 1973. **132**(1): p. 83-93.
145. Perea Velez, M., et al., *Functional analysis of D-alanylation of lipoteichoic acid in the probiotic strain Lactobacillus rhamnosus GG*. Appl Environ Microbiol, 2007. **73**(11): p. 3595-604.
146. Chan, K.G., et al., *Role of D-alanylation of Streptococcus gordonii lipoteichoic acid in innate and adaptive immunity*. Infect Immun, 2007. **75**(6): p. 3033-42.
147. Boman, H.G., *Peptide antibiotics and their role in innate immunity*. Annu Rev Immunol, 1995. **13**: p. 61-92.
148. Poyart, C., et al., *Genetic basis of antibiotic resistance in Streptococcus agalactiae strains isolated in a French hospital*. Antimicrob. Agents Chemother., 2003. **47**(2): p. 794-797.
149. Kiriukhin, M.Y. and F.C. Neuhaus, *D-alanylation of lipoteichoic acid: role of the D-alanyl carrier protein in acylation*. J Bacteriol, 2001. **183**(6): p. 2051-8.
150. Auperin, T.C., et al., *Crystal structure of the N-terminal domain of the group B streptococcus alpha C protein*. J Biol Chem, 2005. **280**(18): p. 18245-52.
151. Stalhammar-Carlemalm, M., L. Stenberg, and G. Lindahl, *Protein Rib: a novel group B streptococcal cell surface protein that confers protective immunity and is expressed by most strains causing invasive infections*. J. Exp. Med., 1993. **177**(6): p. 1593-1603.
152. Nagano, N., et al., *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, 2006. **152**(Pt 3): p. 771-8.
153. Rozhdestvenskaya, A.S., A.A. Totolian, and A.V. Dmitriev, *Inactivation of DNA-binding response regulator Sak189 abrogates beta-antigen expression and affects virulence of Streptococcus agalactiae*. PLoS One, 2010. **5**(4): p. e10212.
154. Spellerberg, B., et al., *Lmb, a protein with similarities to the LraI adhesin family, mediates attachment of Streptococcus agalactiae to human laminin*. Infect Immun, 1999. **67**(2): p. 871-878.
155. Tenenbaum, T., et al., *Streptococcus agalactiae invasion of human brain microvascular endothelial cells is promoted by the laminin-binding protein Lmb*. Microbes Infect, 2007. **9**(6): p. 714-20.

156. Franken, C., et al., *Horizontal gene transfer and host specificity of beta-haemolytic streptococci: the role of a putative composite transposon containing scpB and lmb*. Mol Microbiol, 2001. **41**(4): p. 925-35.
157. Brown, C.K., et al., *Structure of the streptococcal cell wall C5a peptidase*. Proc Natl Acad Sci U S A, 2005. **102**(51): p. 18391-6.
158. Beckmann, C., et al., *Identification of novel adhesins from Group B streptococci by use of phage display reveals that C5a peptidase mediates fibronectin binding*. Infect Immun, 2002. **70**(6): p. 2869-76.
159. Al Safadi, R., et al., *Enhanced expression of lmb gene encoding laminin-binding protein in Streptococcus agalactiae strains harboring IS1548 in scpB-lmb intergenic region*. PLoS One, 2010. **5**(5): p. e10794.
160. Schubert, A., et al., *A fibrinogen receptor from group B Streptococcus interacts with fibrinogen by repetitive units with novel ligand binding sites*. Mol. Microbiol., 2002. **46**(2): p. 557-569.
161. Gutekunst, H., B.J. Eikmanns, and D.J. Reinscheid, *The novel fibrinogen-binding protein FbsB promotes Streptococcus agalactiae invasion into epithelial cells*. Infect Immun, 2004. **72**(6): p. 3495-504.
162. Jacobsson, K., *A novel family of fibrinogen-binding proteins in Streptococcus agalactiae*. Vet Microbiol, 2003. **96**(1): p. 103-13.
163. Rosenau, A., et al., *Evaluation of the ability of Streptococcus agalactiae strains isolated from genital and neonatal specimens to bind to human fibrinogen and correlation with characteristics of the fbsA and fbsB genes*. Infect Immun, 2007. **75**(3): p. 1310-7.
164. Gutekunst, H., B.J. Eikmanns, and D.J. Reinscheid, *Analysis of RogB-controlled virulence mechanisms and gene repression in Streptococcus agalactiae*. Infect Immun, 2003. **71**(9): p. 5056-64.
165. Nobbs, A.H., R.J. Lamont, and H.F. Jenkinson, *Streptococcus adherence and colonization*. Microbiol Mol Biol Rev, 2009. **73**(3): p. 407-50, Table of Contents.
166. Urano-Tashiro, Y., et al., *Binding of the Streptococcus gordonii DLI surface protein Hsa to the host cell membrane glycoproteins CD11b, CD43, and CD50*. Infect Immun, 2008. **76**(10): p. 4686-91.



167. Kinder, S.A. and S.C. Holt, *Localization of the Fusobacterium nucleatum T18 adhesin activity mediating coaggregation with Porphyromonas gingivalis T22*. J Bacteriol, 1993. **175**(3): p. 840-50.
168. Jakubovics, N.S., et al., *Multiple adhesin proteins on the cell surface of Streptococcus gordonii are involved in adhesion to human fibronectin*. Microbiology, 2009. **155**(Pt 11): p. 3572-80.
169. Samen, U., et al., *The surface protein Srr-1 of Streptococcus agalactiae binds human keratin 4 and promotes adherence to epithelial HEp-2 cells*. Infect Immun, 2007. **75**(11): p. 5405-14.
170. Mereghetti, L., et al., *Extensive adaptive changes occur in the transcriptome of Streptococcus agalactiae (group B streptococcus) in response to incubation with human blood*. PLoS One, 2008. **3**(9): p. e3143.
171. Rosini, R., et al., *Identification of novel genomic islands coding for antigenic pilus-like structures in Streptococcus agalactiae*. Mol Microbiol, 2006. **61**(1): p. 126-41.
172. Nobbs, A.H., et al., *Sortase A utilizes an ancillary protein anchor for efficient cell wall anchoring of pili in Streptococcus agalactiae*. Infect Immun, 2008. **76**(8): p. 3550-60.
173. Konto-Ghiorghi, Y., et al., *Dual role for pilus in adherence to epithelial cells and biofilm formation in Streptococcus agalactiae*. PLoS Pathog, 2009. **5**(5): p. e1000422.
174. Maisey, H.C., et al., *Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells*. J Bacteriol, 2007. **189**(4): p. 1464-7.
175. Pezzicoli, A., et al., *Pilus backbone contributes to group B Streptococcus paracellular translocation through epithelial cells*. J Infect Dis, 2008. **198**(6): p. 890-8.
176. Rinaudo, C.D., et al., *Specific involvement of pilus type 2a in biofilm formation in group B Streptococcus*. PLoS One, 2010. **5**(2): p. e9216.
177. Papasergi, S., et al., *The GBS PI-2a pilus is required for virulence in mice neonates*. PLoS One, 2011. **6**(4): p. e18747.
178. Maisey, H.C., et al., *A group B streptococcal pilus protein promotes phagocyte resistance and systemic virulence*. Faseb J, 2008. **22**(6): p. 1715-24.
179. Samen, U., et al., *Rga is a regulator of adherence and pilus formation in Streptococcus agalactiae*. Microbiology, 2011. **157**(Pt 8): p. 2319-27.

180. Tettelin, H., et al., *Towards a universal group B Streptococcus vaccine using multistrain genome analysis*. *Expert. Rev. Vaccines.*, 2006. **5**(5): p. 687-694.
181. Santi, I., et al., *BibA: a novel immunogenic bacterial adhesin contributing to group B Streptococcus survival in human blood*. *Mol Microbiol*, 2007. **63**(3): p. 754-67.
182. Lamy, M.C., et al., *Rapid detection of the "highly virulent" group B Streptococcus ST-17 clone*. *Microbes Infect*, 2006. **8**(7): p. 1714-22.
183. Lamy, M.C., et al., *CovS/CovR of group B streptococcus: a two-component global regulatory system involved in virulence*. *Mol Microbiol*, 2004. **54**(5): p. 1250-68.
184. Spellerberg, B., et al., *Identification of genetic determinants for the hemolytic activity of Streptococcus agalactiae by ISS1 transposition*. *J Bacteriol*, 1999. **181**(10): p. 3212-9.
185. Forquin, M.P., et al., *The putative glycosyltransferase-encoding gene cylJ and the group B Streptococcus (GBS)-specific gene cylK modulate hemolysin production and virulence of GBS*. *Infect Immun*, 2007. **75**(4): p. 2063-6.
186. Liu, G.Y., et al., *Sword and shield: linked group B streptococcal beta-hemolysin/cytolysin and carotenoid pigment function to subvert host phagocyte defense*. *Proc Natl Acad Sci U S A*, 2004. **101**(40): p. 14491-6.
187. Doran, K.S., G.Y. Liu, and V. Nizet, *Group B streptococcal beta-hemolysin/cytolysin activates neutrophil signaling pathways in brain endothelium and contributes to development of meningitis*. *J Clin Invest*, 2003. **112**(5): p. 736-44.
188. Spellerberg, B., et al., *The cyl genes of Streptococcus agalactiae are involved in the production of pigment*. *FEMS Microbiol Lett*, 2000. **188**(2): p. 125-8.
189. Jiang, S.M., et al., *Regulation of virulence by a two-component system in group B streptococcus*. *J Bacteriol*, 2005. **187**(3): p. 1105-13.
190. Rajagopal, L., et al., *Regulation of cytotoxin expression by converging eukaryotic-type and two-component signalling mechanisms in Streptococcus agalactiae*. *Mol Microbiol*, 2006. **62**(4): p. 941-57.
191. Marijuan, P.C., J. Navarro, and R. del Moral, *On prokaryotic intelligence: strategies for sensing the environment*. *Biosystems*, 2009. **99**(2): p. 94-103.
192. Casino, P., V. Rubio, and A. Marina, *The mechanism of signal transduction by two-component systems*. *Curr Opin Struct Biol*, 2010. **20**(6): p. 763-71.
193. Bolotin, A., et al., *The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403*. *Genome Res*, 2001. **11**(5): p. 731-53.

194. Bolotin, A., et al., *Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus*. Nat Biotechnol, 2004. **22**(12): p. 1554-8.
195. Ferretti, J.J., et al., *Complete genome sequence of an M1 strain of Streptococcus pyogenes*. Proc Natl Acad Sci U S A, 2001. **98**(8): p. 4658-63.
196. Tettelin, H., et al., *Complete genome sequence of a virulent isolate of Streptococcus pneumoniae*. Science, 2001. **293**(5529): p. 498-506.
197. Lin, W.J., et al., *Threonine phosphorylation prevents promoter DNA binding of the Group B Streptococcus response regulator CovR*. Mol Microbiol, 2009. **71**(6): p. 1477-95.
198. Levin, J.C. and M.R. Wessels, *Identification of csrR/csrS, a genetic locus that regulates hyaluronic acid capsule synthesis in group A Streptococcus*. Mol Microbiol, 1998. **30**(1): p. 209-19.
199. Engleberg, N.C., et al., *Spontaneous mutations in the CsrRS two-component regulatory system of Streptococcus pyogenes result in enhanced virulence in a murine model of skin and soft tissue infection*. J Infect Dis, 2001. **183**(7): p. 1043-54.
200. Engleberg, N.C., et al., *Contribution of CsrR-regulated virulence factors to the progress and outcome of murine skin infections by Streptococcus pyogenes*. Infect Immun, 2004. **72**(2): p. 623-8.
201. Ikebe, T., et al., *Highly frequent mutations in negative regulators of multiple virulence genes in group A streptococcal toxic shock syndrome isolates*. PLoS Pathog, 2010. **6**(4): p. e1000832.
202. Shea, P.R., et al., *Distinct signatures of diversifying selection revealed by genome analysis of respiratory tract and invasive bacterial populations*. Proc Natl Acad Sci U S A, 2011. **108**(12): p. 5039-44.
203. Graham, M.R., et al., *Virulence control in group A Streptococcus by a two-component gene regulatory system: global expression profiling and in vivo infection modeling*. Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13855-60.
204. Miller, A.A., N.C. Engleberg, and V.J. DiRita, *Repression of virulence genes by phosphorylation-dependent oligomerization of CsrR at target promoters in S. pyogenes*. Mol Microbiol, 2001. **40**(4): p. 976-90.
205. Gryllos, I., J.C. Levin, and M.R. Wessels, *The CsrR/CsrS two-component system of group A Streptococcus responds to environmental Mg<sup>2+</sup>*. Proc Natl Acad Sci U S A, 2003. **100**(7): p. 4227-32.

206. Gryllos, I., et al., *Mg(2+) signalling defines the group A streptococcal CsrRS (CovRS) regulon*. Mol Microbiol, 2007. **65**(3): p. 671-83.
207. Guenzi, E., et al., *A two-component signal-transducing system is involved in competence and penicillin susceptibility in laboratory mutants of Streptococcus pneumoniae*. Mol Microbiol, 1994. **12**(3): p. 505-15.
208. Mascher, T., et al., *The Streptococcus pneumoniae cia regulon: CiaR target sites and transcription profile analysis*. J Bacteriol, 2003. **185**(1): p. 60-70.
209. Sebert, M.E., et al., *Microarray-based identification of htrA, a Streptococcus pneumoniae gene that is regulated by the CiaRH two-component system and contributes to nasopharyngeal colonization*. Infect Immun, 2002. **70**(8): p. 4059-67.
210. Dagkessamanskaia, A., et al., *Interconnection of competence, stress and CiaR regulons in Streptococcus pneumoniae: competence triggers stationary phase autolysis of ciaR mutant cells*. Mol Microbiol, 2004. **51**(4): p. 1071-86.
211. Halfmann, A., et al., *Identification of the genes directly controlled by the response regulator CiaR in Streptococcus pneumoniae: five out of 15 promoters drive expression of small non-coding RNAs*. Mol Microbiol, 2007. **66**(1): p. 110-26.
212. Wu, C., et al., *Regulation of ciaXRH operon expression and identification of the CiaR regulon in Streptococcus mutans*. J Bacteriol, 2010. **192**(18): p. 4669-79.
213. Liu, Y. and R.A. Burne, *Multiple two-component systems of Streptococcus mutans regulate agmatine deiminase gene expression and stress tolerance*. J Bacteriol, 2009. **191**(23): p. 7363-6.
214. He, X., et al., *The cia operon of Streptococcus mutans encodes a unique component required for calcium-mediated autoregulation*. Mol Microbiol, 2008. **70**(1): p. 112-26.
215. Li, J., et al., *The two-component regulatory system CiaRH contributes to the virulence of Streptococcus suis 2*. Vet Microbiol, 2010. **148**(1): p. 99-104.
216. Quach, D., et al., *The CiaR response regulator in group B Streptococcus promotes intracellular survival and resistance to innate immune defenses*. J Bacteriol, 2009. **191**(7): p. 2023-32.
217. Marx, P., et al., *Identification of genes for small non-coding RNAs that belong to the regulon of the two-component regulatory system CiaRH in Streptococcus*. BMC Genomics, 2010. **11**: p. 661.
218. Spellerberg, B., et al., *rgf encodes a novel two-component signal transduction system of Streptococcus agalactiae*. Infect Immun, 2002. **70**(5): p. 2434-40.

219. Safadi, R.A., et al., *Two-component system RgfA/C activates the fbsB gene encoding major fibrinogen-binding protein in highly virulent CC17 clone group B Streptococcus*. PLoS One, 2011. **6**(2): p. e14658.
220. Samen, U.M., B.J. Eikmanns, and D.J. Reinscheid, *The transcriptional regulator RovS controls the attachment of Streptococcus agalactiae to human epithelial cells and the expression of virulence genes*. Infect Immun, 2006. **74**(10): p. 5625-35.
221. Yellaboina, S., et al., *PredictRegulon: a web server for the prediction of the regulatory protein binding sites and operons in prokaryote genomes*. Nucleic Acids Res, 2004. **32**(Web Server issue): p. W318-20.
222. Lauer, P., et al., *Genome analysis reveals pili in Group B Streptococcus*. Science, 2005. **309**(5731): p. 105.
223. Mereghetti, L., et al., *Remodeling of the Streptococcus agalactiae transcriptome in response to growth temperature*. PLoS One, 2008. **3**(7): p. e2785.
224. Adama van Scheltema, P.N., et al., *Biochemical composition of fluids for amniocentesis during fetoscopy*. Gynecol Obstet Invest, 2008. **66**(4): p. 227-30.
225. Mesavage, W.C., et al., *Amino acids in amniotic fluid in the second trimester of gestation*. Pediatr Res, 1985. **19**(10): p. 1021-4.
226. Cho, C.K., et al., *Proteomics analysis of human amniotic fluid*. Mol Cell Proteomics, 2007. **6**(8): p. 1406-15.
227. Shelburne, S.A., 3rd, et al., *Central role of a bacterial two-component gene regulatory system of previously unknown function in pathogen persistence in human saliva*. Proc Natl Acad Sci U S A, 2005. **102**(44): p. 16037-42.
228. Sitkiewicz, I., et al., *Transcriptome adaptation of group B Streptococcus to growth in human amniotic fluid*. PLoS One, 2009. **4**(7): p. e6114.
229. Sitkiewicz, I. and J.M. Musser, *Analysis of growth-phase regulated genes in Streptococcus agalactiae by global transcript profiling*. BMC Microbiol, 2009. **9**: p. 32.
230. Jiang, S.M., et al., *Variation in the group B Streptococcus CsrRS regulon and effects on pathogenicity*. J Bacteriol, 2008. **190**(6): p. 1956-65.
231. Santi, I., et al., *CsrRS regulates group B Streptococcus virulence gene expression in response to environmental pH: a new perspective on vaccine development*. J Bacteriol, 2009. **191**(17): p. 5387-97.

232. Bryan, J.D., et al., *Global transcriptional profiling reveals Streptococcus agalactiae genes controlled by the MtaR transcription factor*. BMC Genomics, 2008. **9**: p. 607.
233. Tettelin, H., et al., *Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae*. Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12391-6.
234. Tettelin, H., et al., *Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome"*. Proc. Natl. Acad. Sci. USA, 2005. **102**(39): p. 13950-13955.
235. Haenni, M., et al., *Diversity and mobility of integrative and conjugative elements in bovine isolates of Streptococcus agalactiae, S. dysgalactiae subsp. dysgalactiae, and S. uberis*. Appl Environ Microbiol, 2010. **76**(24): p. 7957-65.
236. Brochet, M., et al., *Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of Streptococcus agalactiae*. Proc Natl Acad Sci U S A, 2008. **105**(41): p. 15961-6.
237. Toussaint, A. and C. Merlin, *Mobile elements as a combination of functional modules*. Plasmid, 2002. **47**(1): p. 26-35.
238. Burrus, V., et al., *The ICES<sub>St1</sub> element of Streptococcus thermophilus belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration*. Plasmid, 2002. **48**(2): p. 77-97.
239. Wang, H. and P. Mullany, *The large resolvase TndX is required and sufficient for integration and excision of derivatives of the novel conjugative transposon Tn5397*. J Bacteriol, 2000. **182**(23): p. 6577-83.
240. Lewis, J.A. and G.F. Hatfull, *Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins*. Nucleic Acids Res, 2001. **29**(11): p. 2205-16.
241. Nagy, Z. and M. Chandler, *Regulation of transposition in bacteria*. Res Microbiol, 2004. **155**(5): p. 387-98.
242. Celli, J. and P. Trieu-Cuot, *Circularization of Tn916 is required for expression of the transposon-encoded transfer functions: characterization of long tetracycline-inducible transcripts reading through the attachment site*. Mol Microbiol, 1998. **28**(1): p. 103-17.
243. Ishiwa, A. and T. Komano, *PilV adhesins of plasmid R64 thin pili specifically bind to the lipopolysaccharides of recipient cells*. J Mol Biol, 2004. **343**(3): p. 615-25.

244. Funnell, B.E.a.G., J. P. , *Plasmid Biology*. 2004.
245. Babic, A., et al., *Direct visualization of horizontal gene transfer*. *Science*, 2008. **319**(5869): p. 1533-6.
246. Shu, A.C., et al., *Evidence of DNA transfer through F-pilus channels during Escherichia coli conjugation*. *Langmuir*, 2008. **24**(13): p. 6796-802.
247. Anderson, D.G. and L.L. McKay, *Genetic and physical characterization of recombinant plasmids associated with cell aggregation and high-frequency conjugal transfer in Streptococcus lactis ML3*. *J Bacteriol*, 1984. **158**(3): p. 954-62.
248. Grohmann, E., G. Muth, and M. Espinosa, *Conjugative plasmid transfer in gram-positive bacteria*. *Microbiol. Mol. Biol. Rev.*, 2003. **67**(2): p. 277-301.
249. Waters, C.M., et al., *An amino-terminal domain of Enterococcus faecalis aggregation substance is required for aggregation, bacterial internalization by epithelial cells and binding to lipoteichoic acid*. *Mol Microbiol*, 2004. **52**(4): p. 1159-71.
250. Alvarez-Martinez, C.E. and P.J. Christie, *Biological diversity of prokaryotic type IV secretion systems*. *Microbiol Mol Biol Rev*, 2009. **73**(4): p. 775-808.
251. Talay, S.R., et al., *Domain structure and conserved epitopes of Sfb protein, the fibronectin-binding adhesin of Streptococcus pyogenes*. *Mol Microbiol*, 1994. **13**(3): p. 531-9.
252. Navarre, W.W. and O. Schneewind, *Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope*. *Microbiol Mol Biol Rev*, 1999. **63**(1): p. 174-229.
253. Munro, G.H., et al., *A protein fragment of streptococcal cell surface antigen I/II which prevents adhesion of Streptococcus mutans*. *Infect Immun*, 1993. **61**(11): p. 4590-8.
254. Brady, L.J., et al., *Differentiation of salivary agglutinin-mediated adherence and aggregation of mutans streptococci by use of monoclonal antibodies against the major surface adhesin PI*. *Infect Immun*, 1992. **60**(3): p. 1008-17.
255. Daep, C.A., et al., *Structural characterization of peptide-mediated inhibition of Porphyromonas gingivalis biofilm formation*. *Infect Immun*, 2006. **74**(10): p. 5756-62.
256. Brady, L.J., et al., *The changing faces of Streptococcus antigen I/II polypeptide family adhesins*. *Mol Microbiol*, 2010. **77**(2): p. 276-86.
257. Srinivas, P., A.O. Kilic, and M.N. Vijayakumar, *Site-specific nicking in vitro at ori T by the DNA relaxase of Tn5252*. *Plasmid*, 1997. **37**(1): p. 42-50.

258. Lucas, M., et al., *Relaxase DNA binding and cleavage are two distinguishable steps in conjugative DNA processing that involve different sequence elements of the nic site*. J Biol Chem, 2010. **285**(12): p. 8918-26.
259. Garcillan-Barcia, M.P., M.V. Francia, and F. de la Cruz, *The diversity of conjugative relaxases and its application in plasmid classification*. FEMS Microbiol Rev, 2009. **33**(3): p. 657-87.
260. Byrd, D.R. and S.W. Matson, *Nicking by transesterification: the reaction catalysed by a relaxase*. Mol Microbiol, 1997. **25**(6): p. 1011-22.
261. Grandoso, G., et al., *Two active-site tyrosyl residues of protein TrwC act sequentially at the origin of transfer during plasmid R388 conjugation*. J Mol Biol, 2000. **295**(5): p. 1163-72.
262. Lee, C.A. and A.D. Grossman, *Identification of the origin of transfer (oriT) and DNA relaxase required for conjugation of the integrative and conjugative element ICEBsI of Bacillus subtilis*. J Bacteriol, 2007. **189**(20): p. 7254-61.
263. Rocco, J.M. and G. Churchward, *The integrase of the conjugative transposon Tn916 directs strand- and sequence-specific cleavage of the origin of conjugal transfer, oriT, by the endonuclease Orf20*. J Bacteriol, 2006. **188**(6): p. 2207-13.
264. Chen, Y., J.H. Staddon, and G.M. Dunny, *Specificity determinants of conjugative DNA processing in the Enterococcus faecalis plasmid pCF10 and the Lactococcus lactis plasmid pRS01*. Mol Microbiol, 2007. **63**(5): p. 1549-64.
265. Llosa, M., et al., *Bacterial conjugation: a two-step mechanism for DNA transport*. Mol Microbiol, 2002. **45**(1): p. 1-8.
266. Matson, S.W., J.K. Sampson, and D.R. Byrd, *F plasmid conjugative DNA transfer: the TraI helicase activity is essential for DNA strand transfer*. J Biol Chem, 2001. **276**(4): p. 2372-9.
267. Jaworski, D.D. and D.B. Clewell, *A functional origin of transfer (oriT) on the conjugative transposon Tn916*. J Bacteriol, 1995. **177**(22): p. 6644-51.
268. Iyer, L.M., et al., *Evolutionary history and higher order classification of AAA+ ATPases*. J Struct Biol, 2004. **146**(1-2): p. 11-31.
269. Chen, Y., et al., *Enterococcus faecalis PcfC, a spatially localized substrate receptor for type IV secretion of the pCF10 transfer intermediate*. J Bacteriol, 2008. **190**(10): p. 3632-45.



270. Abajy, M.Y., et al., *A type IV-secretion-like system is required for conjugative DNA transport of broad-host-range plasmid pIP501 in gram-positive bacteria*. J Bacteriol, 2007. **189**(6): p. 2487-96.
271. Steen, J.A., et al., *The putative coupling protein TcpA interacts with other pCW3-encoded proteins to form an essential part of the conjugation complex*. J Bacteriol, 2009. **191**(9): p. 2926-33.
272. Berger, B.R. and P.J. Christie, *Genetic complementation analysis of the Agrobacterium tumefaciens virB operon: virB2 through virB11 are essential virulence genes*. J Bacteriol, 1994. **176**(12): p. 3646-60.
273. Smillie, C., et al., *Mobility of plasmids*. Microbiol Mol Biol Rev, 2010. **74**(3): p. 434-52.
274. Wozniak, R.A. and M.K. Waldor, *Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow*. Nat Rev Microbiol, 2010. **8**(8): p. 552-63.
275. Storrs, M.J., et al., *Conjugative transposition of Tn916 requires the excisive and integrative activities of the transposon-encoded integrase*. J Bacteriol, 1991. **173**(14): p. 4347-52.
276. Dimopoulou, I.D., et al., *Molecular epidemiology of unrelated clusters of multiresistant strains of Haemophilus influenzae*. J Infect Dis, 1992. **165**(6): p. 1069-75.
277. Dimopoulou, I.D., et al., *A molecular analysis of Greek and UK Haemophilus influenzae conjugative resistance plasmids*. J Antimicrob Chemother, 1997. **39**(3): p. 303-7.
278. Leaves, N.I., et al., *Epidemiological studies of large resistance plasmids in Haemophilus*. J Antimicrob Chemother, 2000. **45**(5): p. 599-604.
279. Schleper, C., et al., *A multicopy plasmid of the extremely thermophilic archaeon Sulfolobus effects its transfer to recipients by mating*. J Bacteriol, 1995. **177**(15): p. 4417-26.
280. She, Q., B. Shen, and L. Chen, *Archaeal integrases and mechanisms of gene capture*. Biochem Soc Trans, 2004. **32**(Pt 2): p. 222-6.
281. Parker, C. and R. Meyer, *Mechanisms of strand replacement synthesis for plasmid DNA transferred by conjugation*. J Bacteriol, 2005. **187**(10): p. 3400-6.

282. Rees, C.E. and B.M. Wilkins, *Protein transfer into the recipient cell during bacterial conjugation: studies with F and RP4*. Mol Microbiol, 1990. **4**(7): p. 1199-205.
283. Wilkins, B.M. and A.T. Thomas, *DNA-independent transport of plasmid primase protein between bacteria by the HII conjugation system*. Mol Microbiol, 2000. **38**(3): p. 650-7.
284. Hagege, J., et al., *Identification of a gene encoding the replication initiator protein of the Streptomyces integrating element, pSAM2*. Plasmid, 1994. **31**(2): p. 166-83.
285. Possoz, C., et al., *The integrative element pSAM2 from Streptomyces: kinetics and mode of conjugal transfer*. Mol Microbiol, 2001. **42**(1): p. 159-66.
286. Ike, Y., et al., *Efficient transfer of the pheromone-independent Enterococcus faecium plasmid pMG1 (Gmr) (65.1 kilobases) to Enterococcus strains during broth mating*. J Bacteriol, 1998. **180**(18): p. 4886-92.
287. Clewell, D.B., et al., *Characterization of three plasmid deoxyribonucleic acid molecules in a strain of Streptococcus faecalis: identification of a plasmid determining erythromycin resistance*. J Bacteriol, 1974. **117**(1): p. 283-9.
288. El-Solh, N., et al., *Molecular studies and possible relatedness between R plasmids from groups B and D streptococci*. Antimicrob Agents Chemother, 1978. **14**(1): p. 19-23.
289. Godon, J.J., et al., *Molecular analysis of the Lactococcus lactis sex factor*. Dev Biol Stand, 1995. **85**: p. 423-30.
290. Clewell, D.B., et al., *The conjugative transposon Tn916 of Enterococcus faecalis: structural analysis and some key factors involved in movement*. Dev Biol Stand, 1995. **85**: p. 11-7.
291. Manning, P.A., L. Beutin, and M. Achtman, *Outer membrane of Escherichia coli: properties of the F sex factor traT protein which is involved in surface exclusion*. J Bacteriol, 1980. **142**(1): p. 285-94.
292. Harrison, J.L., et al., *Surface exclusion specificity of the TraT lipoprotein is determined by single alterations in a five-amino-acid region of the protein*. Mol Microbiol, 1992. **6**(19): p. 2825-32.
293. Minkley, E.G., Jr. and K. Ippen-Ihler, *Identification of a membrane protein associated with expression of the surface exclusion region of the F transfer operon*. J Bacteriol, 1977. **129**(3): p. 1613-22.

294. Audette, G.F., et al., *Entry exclusion in F-like plasmids requires intact TraG in the donor that recognizes its cognate TraS in the recipient*. *Microbiology*, 2007. **153**(Pt 2): p. 442-51.
295. Anthony, K.G., et al., *Comparison of proteins involved in pilus synthesis and mating pair stabilization from the related plasmids F and R100-1: insights into the mechanism of conjugation*. *J Bacteriol*, 1999. **181**(17): p. 5149-59.
296. Maxwell, A., R. Craigie, and K. Mizuuchi, *B protein of bacteriophage mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer*. *Proc Natl Acad Sci U S A*, 1987. **84**(3): p. 699-703.
297. Adzuma, K. and K. Mizuuchi, *Target immunity of Mu transposition reflects a differential distribution of Mu B protein*. *Cell*, 1988. **53**(2): p. 257-66.
298. Stellwagen, A.E. and N.L. Craig, *Gain-of-function mutations in TnsC, an ATP-dependent transposition protein that activates the bacterial transposon Tn7*. *Genetics*, 1997. **145**(3): p. 573-85.
299. Stellwagen, A.E. and N.L. Craig, *Avoiding self: two Tn7-encoded proteins mediate target immunity in Tn7 transposition*. *Embo J*, 1997. **16**(22): p. 6823-34.
300. Skelding, Z., J. Queen-Baker, and N.L. Craig, *Alternative interactions between the Tn7 transposase and the Tn7 target DNA binding protein regulate target immunity and transposition*. *Embo J*, 2003. **22**(21): p. 5904-17.
301. Birge E. A., *Bacterial and bacteriophage genetics*. Fourth Edition ed. 2000, New York. 255,263,271,277.
302. Madsen, P.L., et al., *The genetic switch regulating activity of early promoters of the temperate lactococcal bacteriophage TP901-1*. *J Bacteriol*, 1999. **181**(24): p. 7430-8.
303. Auchtung, J.M., et al., *Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICEBs1 of Bacillus subtilis*. *Mol Microbiol*, 2007. **64**(6): p. 1515-28.
304. Auchtung, J.M., et al., *Regulation of a Bacillus subtilis mobile genetic element by intercellular signaling and the global DNA damage response*. *Proc Natl Acad Sci U S A*, 2005. **102**(35): p. 12554-9.
305. Bose, B., et al., *A conserved anti-repressor controls horizontal gene transfer by proteolysis*. *Mol Microbiol*, 2008. **70**(3): p. 570-82.

306. Celli, J., C. Poyart, and P. Trieu-Cuot, *Use of an excision reporter plasmid to study the intracellular mobility of the conjugative transposon Tn916 in gram-positive bacteria.* Microbiology, 1997. **143 ( Pt 4)**: p. 1253-61.
307. Whittle, G., N.B. Shoemaker, and A.A. Salyers, *Characterization of genes involved in modulation of conjugal transfer of the Bacteroides conjugative transposon CTnDOT.* J Bacteriol, 2002. **184(14)**: p. 3839-47.
308. Wang, Y., et al., *Translational control of tetracycline resistance and conjugation in the Bacteroides conjugative transposon CTnDOT.* J Bacteriol, 2005. **187(8)**: p. 2673-80.
309. Minoia, M., et al., *Stochasticity and bistability in horizontal transfer control of a genomic island in Pseudomonas.* Proc Natl Acad Sci U S A, 2008. **105(52)**: p. 20792-7.
310. Sentschilo, V., et al., *Unusual integrase gene expression on the clc genomic island in Pseudomonas sp. strain B13.* J Bacteriol, 2003. **185(15)**: p. 4530-8.
311. Sentschilo, V., A.J. Zehnder, and J.R. van der Meer, *Characterization of two alternative promoters for integrase expression in the clc genomic island of Pseudomonas sp. strain B13.* Mol Microbiol, 2003. **49(1)**: p. 93-104.
312. Beaber, J.W., B. Hochhut, and M.K. Waldor, *SOS response promotes horizontal dissemination of antibiotic resistance genes.* Nature, 2004. **427(6969)**: p. 72-4.
313. Beaber, J.W. and M.K. Waldor, *Identification of operators and promoters that control SXT conjugative transfer.* J Bacteriol, 2004. **186(17)**: p. 5945-9.
314. Wozniak, R.A. and M.K. Waldor, *A toxin-antitoxin system promotes the maintenance of an integrative conjugative element.* PLoS Genet, 2009. **5(3)**: p. e1000439.
315. Dunny, G.M., B.L. Brown, and D.B. Clewell, *Induced cell aggregation and mating in Streptococcus faecalis: evidence for a bacterial sex pheromone.* Proc Natl Acad Sci U S A, 1978. **75(7)**: p. 3479-83.
316. Chandler, J.R., et al., *Specific control of endogenous cCF10 pheromone by a conserved domain of the pCF10-encoded regulatory protein PrgY in Enterococcus faecalis.* J Bacteriol, 2005. **187(14)**: p. 4830-43.
317. Kozłowicz, B.K., et al., *Molecular basis for control of conjugation by bacterial pheromone and inhibitor peptides.* Mol Microbiol, 2006. **62(4)**: p. 958-69.
318. Weidlich, G., R. Wirth, and D. Galli, *Sex pheromone plasmid pADI-encoded surface exclusion protein of Enterococcus faecalis.* Mol Gen Genet, 1992. **233(1-2)**: p. 161-8.

319. Yagi, Y., et al., *Plasmid content of Streptococcus faecalis strain 39-5 and identification of a pheromone (cPD1)-induced surface antigen*. J Gen Microbiol, 1983. **129**(4): p. 1207-15.
320. Nakayama, J., et al., *The prgQ gene of the Enterococcus faecalis tetracycline resistance plasmid pCF10 encodes a peptide inhibitor, iCF10*. J Bacteriol, 1994. **176**(23): p. 7405-8.
321. Chandler, J.R. and G.M. Dunny, *Characterization of the sequence specificity determinants required for processing and control of sex pheromone by the intramembrane protease Eep and the plasmid-encoded protein PrgY*. J Bacteriol, 2008. **190**(4): p. 1172-83.
322. Wardal, E., E. Sadowy, and W. Hryniewicz, *Complex nature of enterococcal pheromone-responsive plasmids*. Pol J Microbiol, 2010. **59**(2): p. 79-87.
323. Conway, T. and G.K. Schoolnik, *Microarray expression profiling: capturing a genome-wide portrait of the transcriptome*. Mol Microbiol, 2003. **47**(4): p. 879-89.
324. Biswas, I., et al., *High-efficiency gene inactivation and replacement system for gram-positive bacteria*. J. Bacteriol., 1993. **175**(11): p. 3628-3635.
325. Taylor, C.M., et al., *Listeria monocytogenes relA and hpt mutants are impaired in surface-attached growth and virulence*. J Bacteriol, 2002. **184**(3): p. 621-8.
326. Yang, Y.H., et al., *Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation*. Nucleic Acids Res, 2002. **30**(4): p. e15.
327. Reiner, A., D. Yekutieli, and Y. Benjamini, *Identifying differentially expressed genes using false discovery rate controlling procedures*. Bioinformatics, 2003. **19**(3): p. 368-75.
328. de Hoon, M.J., et al., *Open source clustering software*. Bioinformatics, 2004. **20**(9): p. 1453-4.
329. Saldanha, A.J., *Java Treeview--extensible visualization of microarray data*. Bioinformatics, 2004. **20**(17): p. 3246-8.
330. Novichkov, P.S., et al., *RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach*. Nucleic Acids Res, 2010. **38**(Web Server issue): p. W299-307.

331. Ellwood, D.C. and D.W. Tempest, *Influence of culture pH on the content and composition of teichoic acids in the walls of Bacillus subtilis*. J Gen Microbiol, 1972. **73**(2): p. 395-402.
332. McCormick, N.E., S.A. Halperin, and S.F. Lee, *Regulation of D-alanylation of lipoteichoic acid in Streptococcus gordonii*. Microbiology, 2011. **157**(Pt 8): p. 2248-56.
333. Dubrac, S., et al., *A matter of life and death: cell wall homeostasis and the WalkR (YycGF) essential signal transduction pathway*. Mol Microbiol, 2008. **70**(6): p. 1307-22.
334. Sitkiewicz, I. and J.M. Musser, *Expression microarray and mouse virulence analysis of four conserved two-component gene regulatory systems in group a streptococcus*. Infect Immun, 2006. **74**(2): p. 1339-51.
335. Emilianowicz-Czerska, W., et al., *Studies on biosynthesis of polymyxin E (colistin). I. Isolation and characteristics of mucoiless R-mutants of Bacillus polymyxa K-1*. Acta Microbiol Pol B, 1972. **4**(4): p. 211-5.
336. Majchrzykiewicz, J.A., O.P. Kuipers, and J.J. Bijlsma, *Generic and specific adaptive responses of Streptococcus pneumoniae to challenge with three distinct antimicrobial peptides, bacitracin, LL-37, and nisin*. Antimicrob Agents Chemother, 2010. **54**(1): p. 440-51.
337. Pincus, S.H., et al., *Group B streptococcal opacity variants*. J Bacteriol, 1992. **174**(11): p. 3739-49.
338. Fogg, G.C., C.M. Gibson, and M.G. Caparon, *The identification of rofA, a positive-acting regulatory component of prtF expression: use of an m gamma delta-based shuttle mutagenesis strategy in Streptococcus pyogenes*. Mol Microbiol, 1994. **11**(4): p. 671-84.
339. Luo, F., S. Lizano, and D.E. Bessen, *Heterogeneity in the polarity of Nra regulatory effects on streptococcal pilus gene transcription and virulence*. Infect Immun, 2008. **76**(6): p. 2490-7.
340. Kreikemeyer, B., et al., *Group A streptococcal RofA-type global regulators exhibit a strain-specific genomic presence and regulation pattern*. Microbiology, 2002. **148**(Pt 5): p. 1501-11.
341. Lizano, S., et al., *Impact of orthologous gene replacement on the circuitry governing pilus gene transcription in streptococci*. PLoS One, 2008. **3**(10): p. e3450.

342. Herbert, S., et al., *Molecular basis of resistance to muramidase and cationic antimicrobial peptide activity of lysozyme in staphylococci*. PLoS Pathog, 2007. **3**(7): p. e102.
343. Kjos, M., et al., *The abi proteins and their involvement in bacteriocin self-immunity*. J Bacteriol, 2010. **192**(8): p. 2068-76.
344. Yamamoto, Y., et al., *Respiration metabolism of Group B Streptococcus is activated by environmental haem and quinone and contributes to virulence*. Mol Microbiol, 2005. **56**(2): p. 525-34.
345. Yamamoto, Y., et al., *Roles of environmental heme, and menaquinone, in streptococcus agalactiae*. Biometals, 2006. **19**(2): p. 205-10.
346. Yamamoto, Y., et al., *The Group B Streptococcus NADH oxidase Nox-2 is involved in fatty acid biosynthesis during aerobic growth and contributes to virulence*. Mol Microbiol, 2006. **62**(3): p. 772-85.
347. Gyan, S., et al., *Regulatory loop between redox sensing of the NADH/NAD(+) ratio by Rex (YdiH) and oxidation of NADH by NADH dehydrogenase Ndh in Bacillus subtilis*. J Bacteriol, 2006. **188**(20): p. 7062-71.
348. Brinster, S., et al., *Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens*. Nature, 2009. **458**(7234): p. 83-6.
349. Ton-That, H., et al., *Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates*. J Biol Chem, 2000. **275**(13): p. 9876-81.
350. Baccini, A., et al., *Stratégies pour l'analyse statistique de données transcriptomiques*. Journal de la Société Française de Statistique, 2005. **tome146**.
351. Mitrophanov, A.Y. and E.A. Groisman, *Positive feedback in cellular control systems*. Bioessays, 2008. **30**(6): p. 542-55.
352. Ray, J.C. and O.A. Igoshin, *Adaptable functionality of transcriptional feedback in bacterial two-component systems*. PLoS Comput Biol, 2010. **6**(2): p. e1000676.
353. Haydon, D.J. and J.R. Guest, *A new family of bacterial regulatory proteins*. FEMS Microbiol Lett, 1991. **63**(2-3): p. 291-5.
354. Rigali, S., et al., *Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies*. J Biol Chem, 2002. **277**(15): p. 12507-15.

355. Novichkov, P.S., et al., *RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes*. *Nucleic Acids Res*, 2010. **38**(Database issue): p. D111-8.
356. Chen, P.M., et al., *Role of GlnR in acid-mediated repression of genes encoding proteins involved in glutamine and glutamate metabolism in Streptococcus mutans*. *Appl Environ Microbiol*, 2010. **76**(8): p. 2478-86.
357. Brown, S.W. and A.L. Sonenshein, *Autogenous regulation of the Bacillus subtilis glnRA operon*. *J Bacteriol*, 1996. **178**(8): p. 2450-4.
358. Doroshchuk, N.A., M.S. Gel'fand, and D.A. Rodionov, *[Regulation of nitrogen metabolism in gram-positive bacteria]*. *Mol Biol (Mosk)*, 2006. **40**(5): p. 919-26.
359. Schulz, A. and W. Schumann, *hrcA, the first gene of the Bacillus subtilis dnaK operon encodes a negative regulator of class I heat shock genes*. *J Bacteriol*, 1996. **178**(4): p. 1088-93.
360. Hecker, M., W. Schumann, and U. Volker, *Heat-shock and general stress response in Bacillus subtilis*. *Mol Microbiol*, 1996. **19**(3): p. 417-28.
361. Narberhaus, F., *Negative regulation of bacterial heat shock genes*. *Mol Microbiol*, 1999. **31**(1): p. 1-8.
362. Servant, P. and P. Mazodier, *Negative regulation of the heat shock response in Streptomyces*. *Arch Microbiol*, 2001. **176**(4): p. 237-42.
363. Kim, S.N., et al., *Molecular cloning and characterization of groESL operon in Streptococcus pneumoniae*. *Mol Cells*, 2001. **11**(3): p. 360-8.
364. Kwon, H.Y., et al., *Reduction-sensitive and cysteine residue-mediated Streptococcus pneumoniae HrcA oligomerization in vitro*. *Mol Cells*, 2009. **27**(2): p. 149-57.
365. Li, J.S., et al., *Transcriptome Analysis of Adaptive Heat Shock Response of Streptococcus thermophilus*. *PLoS One*, 2011. **6**(10): p. e25777.
366. Jayaraman, G.C., J.E. Penders, and R.A. Burne, *Transcriptional analysis of the Streptococcus mutans hrcA, grpE and dnaK genes and regulation of expression in response to heat shock and environmental acidification*. *Mol Microbiol*, 1997. **25**(2): p. 329-41.
367. Lemos, J.A., Y.Y. Chen, and R.A. Burne, *Genetic and physiologic analysis of the groE operon and role of the HrcA repressor in stress gene regulation and acid tolerance in Streptococcus mutans*. *J Bacteriol*, 2001. **183**(20): p. 6074-84.



368. Woodbury, R. and W.G. Haldenwang, *HrcA is a negative regulator of the dnaK and groESL operons of Streptococcus pyogenes*. Biochem Biophys Res Commun, 2003. **302**(4): p. 722-7.
369. Pagels, M., et al., *Redox sensing by a Rex-family repressor is involved in the regulation of anaerobic gene expression in Staphylococcus aureus*. Mol Microbiol, 2010. **76**(5): p. 1142-61.
370. Kinch, L.N., et al., *Identification of novel restriction endonuclease-like fold families among hypothetical proteins*. Nucleic Acids Res, 2005. **33**(11): p. 3598-605.
371. Desai, B.V. and D.A. Morrison, *An unstable competence-induced protein, CoiA, promotes processing of donor DNA after uptake during genetic transformation in Streptococcus pneumoniae*. J Bacteriol, 2006. **188**(14): p. 5177-86.
372. Zhao, B. and W.A. Houry, *Acid stress response in enteropathogenic gamma-proteobacteria: an aptitude for survival*. Biochem Cell Biol, 2010. **88**(2): p. 301-14.
373. Marquis, R.E., et al., *Arginine deiminase system and bacterial adaptation to acid environments*. Appl Environ Microbiol, 1987. **53**(1): p. 198-200.
374. Poyart, C. and P. Trieu-Cuot, *A broad-host-range mobilizable shuttle vector for the construction of transcriptional fusions to beta-galactosidase in gram-positive bacteria*. FEMS Microbiol Lett, 1997. **156**(2): p. 193-8.
375. Ferrari, E., S.M. Howard, and J.A. Hoch, *Effect of stage 0 sporulation mutations on subtilisin expression*. J Bacteriol, 1986. **166**(1): p. 173-9.
376. Khan, E., et al., *Retroviral integrase domains: DNA binding and the recognition of LTR sequences*. Nucleic Acids Res., 1991. **19**(4): p. 851-860.
377. van der Ploeg, J.R., *Analysis of CRISPR in Streptococcus mutans suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages*. Microbiology, 2009. **155**(Pt 6): p. 1966-76.
378. von Heijne, G., *Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule*. J Mol Biol, 1992. **225**(2): p. 487-94.
379. Claros, M.G. and G. von Heijne, *TopPred II: an improved software for membrane protein structure predictions*. Comput Appl Biosci, 1994. **10**(6): p. 685-6.
380. Pichon, C., et al., *An in silico model for identification of small RNAs in whole bacterial genomes: characterization of antisense RNAs in pathogenic Escherichia coli and Streptococcus agalactiae strains*. Nucleic Acids Res, 2011. **2011**: p. 1.

381. Duval-Valentin, G., et al., *Transient promoter formation: a new feedback mechanism for regulation of IS911 transposition*. *Embo J*, 2001. **20**(20): p. 5802-11.
382. Bellanger, X., et al., *Regulation of excision of integrative and potentially conjugative elements from *Streptococcus thermophilus*: role of the arp1 repressor*. *J Mol Microbiol Biotechnol*, 2008. **14**(1-3): p. 16-21.
383. Bellanger, X., et al., *Conjugative transfer of the integrative conjugative elements ICES<sub>t1</sub> and ICES<sub>t3</sub> from *Streptococcus thermophilus**. *J Bacteriol*, 2009. **191**(8): p. 2764-75.
384. Bellanger, X., et al., *Derepression of excision of integrative and potentially conjugative elements from *Streptococcus thermophilus* by DNA damage response: implication of a cI-related repressor*. *J Bacteriol*, 2007. **189**(4): p. 1478-81.
385. Callanan, M.J., T.P. Beresford, and R.P. Ross, *Genetic diversity in the lactose operons of *Lactobacillus helveticus* strains and its relationship to the role of these strains as commercial starter cultures*. *Appl Environ Microbiol*, 2005. **71**(3): p. 1655-8.
386. Dorman, C.J., *Regulatory integration of horizontally-transferred genes in bacteria*. *Front Biosci*, 2009. **14**: p. 4103-12.
387. Fass, E. and E.A. Groisman, *Control of *Salmonella* pathogenicity island-2 gene expression*. *Curr Opin Microbiol*, 2009. **12**(2): p. 199-204.
388. McAdams, H.H., B. Srinivasan, and A.P. Arkin, *The evolution of genetic regulatory systems in bacteria*. *Nat Rev Genet*, 2004. **5**(3): p. 169-78.
389. Lynch, M., *The evolution of genetic networks by non-adaptive processes*. *Nat Rev Genet*, 2007. **8**(10): p. 803-13.
390. Jeong, S., et al., *The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species*. *Cell*, 2008. **132**(5): p. 783-93.
391. Carroll, S.B., *Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution*. *Cell*, 2008. **134**(1): p. 25-36.
392. Smyth, D.S. and D.A. Robinson, *Integrative and sequence characteristics of a novel genetic element, ICE6013, in *Staphylococcus aureus**. *J Bacteriol*, 2009. **191**(19): p. 5964-75.

## **4. ANNEXES**

### **I. Partie analyse des réseaux de régulation**

**Tableau S1: Conservation des systèmes à deux composants au sein de espèces du genre streptocoque**

family	NEM316	COH1	A909	2603V/R	S. pyogenes		S. galloyticus		S. infantarius	S. salivarius	S. mutans UA159	S. suis	S.pneumoni ae R6	S. mitis B6	S. sanguinis SK36	S. uberis 0140J
					M1	UCN34										
OmpR	gbs0121	SAN_0154	SAK_0174	SAG0123	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	SUB0203
	gbs0122	SAN_0155	SAK_0175	SAG0124	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	SUB0202
LyfT	gbs0180	SAN_0221	SAK_0248	SAG0182	NP	NP	NP	NP	NP	SMU.577	NP	NP	NP	NP	NP	NP
	gbs0181	SAN_0222	SAK_0249	SAG0183	NP	NP	NP	NP	NP	SMU.576	NP	NP	NP	NP	NP	NP
LuxR/GerR	gbs0298	SAN_0347	SAK_0380	SAG0310	NP	GALLO_0551	STRINF_0067 6	STRSA0001_ 1126	STRINF_0067 5	SMU.1548c	SSU098_169 4	spr1815	smi_0258	SSA_1973	SSA_1972	NP
	gbs0299	SAN_0348	SAK_0381	SAG0311	NP	GALLO_0552	STRINF_0067 5	STRSA0001_ 1127	STRINF_0067 5	SMU.1547c	SSU098_169 5	spr1814	smi_0259	SSA_1972	SSA_1972	NP
LiaS	gbs0309	SAN_0358	SAK_0391	SAG0321	NP	GALLO_0496	STRINF_0024 8	STRSA0001_ 0023	STRINF_0024 9	SMU.486	SSU098_041 7	spr0343	smi_1742	SSA_1843	SSA_1843	SUB1383
LiaR	gbs0310	SAN_0359	SAK_0392	SAG0322	NP	GALLO_0497	STRINF_0024 9	STRSA0001_ 0022	STRINF_0024 9	SMU.487	SSU098_041 9	spr0344	smi_1741	SSA_1842	SSA_1842	SUB1382
OmpR	gbs0429	SAN_0439	SAK_0467	SAG0393	NP	NP	NP	STRSA0001_ 1978	NP	NP	NP	spr0076	smi_0092	SSA_1113	SSA_1113	NP
	gbs0430	SAN_0440	SAK_0468	SAG0394	NP	NP	NP	STRSA0001_ 1979	NP	NP	NP	spr0077	smi_0093	SSA_1114	SSA_1114	NP
vncR	gbs0597	SAN_0659	SAK_0701	SAG0616	NP	NP	NP	STRSA0001_ 1149	NP	NP	NP	spr0528	smi_0693	SSA_0896	SSA_0896	NP
	gbs0598	SAN_0660	SAK_0702	SAG0617	NP	NP	NP	STRSA0001_ 1150	NP	NP	NP	spr0529	smi_0694	SSA_0897	SSA_0897	NP
OmpR	gbs0685	SAN_0783	SAK_0838	SAG0712	NP	NP	NP	NP	NP	NP	NP	NP	NP	SSA_1575	SSA_1575	NP
	gbs0741	SAN_0790	SAK_0845	SAG0719	NP	GALLO_1554	STRINF_0190 9	STRSA0001_ 1812	STRINF_0190 8	SMU.1517	SSU098_137 6	spr1107	smi_1177	SSA_1565	SSA_1565	SUB0552
waiK	gbs0742	SAN_0791	SAK_0846	SAG0720	NP	GALLO_1553	STRINF_0190 8	STRSA0001_ 1813	STRINF_0190 8	SMU.1516	SSU098_137 5	spr1106	smi_1176	SSA_1564	SSA_1564	SUB0553
	gbs0963	SAN_1088	SAK_1071	SAG0976	NP	GALLO_1088	STRINF_0126 2	STRSA0001_ 1211	STRINF_0126 1	SMU.1008	SSU098_088 8	spr1474	smi_0649	SSA_1685	SSA_1685	SUB0917
graS	gbs0964	SAN_1089	SAK_1072	SAG0977	NP	GALLO_1087	STRINF_0126 1	STRSA0001_ 1212	STRINF_0126 1	SMU.1009	SSU098_088 7	spr1473	smi_0650	SSA_1684	SSA_1684	SUB0916
	gbs1019	SAN_1100	SAK_1079	SAG0984	NP	GALLO_1141	STRINF_0133 5	STRSA0001_ 1605	STRINF_0133 6	SMU.1128	SSU098_110 4	spr0708	smi_0852	SSA_0960	SSA_0960	SUB0957
ciaR	gbs1020	SAN_1101	SAK_1080	SAG0985	NP	GALLO_1142	STRINF_0133 6	STRSA0001_ 1604	STRINF_0133 6	SMU.1129	SSU098_110 5	spr0707	smi_0851	SSA_0959	SSA_0959	SUB0958
	gbs1051	SAN_1141	SAK_1111	SAG1016	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP



**Tableau S2: Liste des gènes codant pour des régulateurs transcriptionnels et des histidines kinases présents dans la classification hiérarchique.**

N°cluster	gene	
16	gbs0094	transcription repressor of class I heat-shock (HrcA)
22	gbs0098	transcriptional regulator (GntR family)
8	gbs0118	transcription repressor of ribose operon
34	gbs0150	transcriptional regulator
27	gbs0168	transcriptional regulator
9	gbs0469	transcriptional regulator (AraC/XylS family)
14	gbs0597	two-component response regulator VncR
27	gbs0618	transcriptional regulator (TetR/AcrR family) N-terminal part
20	gbs0627	transcriptional regulator (N-terminal part)
20	gbs0685	Similar to two-component response regulator
24	gbs0823	transcription factor (CoiA family)
9	gbs0833	transcriptional regulator
2	gbs0928	transcriptional regulator (GntR family)
29	gbs0963	two-component response regulator
29	gbs0964	two-component sensor histidine kinase
28	gbs1019	two-component sensor histidine kinase
28	gbs1020	two-component response regulator
1	gbs1022	phosphate uptake regulatory protein
27	gbs1196	transcriptional regulator
12	gbs1397	two-component sensor histidine kinase
12	gbs1398	two-component response regulator
27	gbs1402	transcriptional regulator
14	gbs1426	transcription regulator RofA related
32	gbs1434	pyrimidine biosynthetic operon repressor
32	gbs1437	transcriptional regulator (LysR/MarR family)
33	gbs1479	RogB transcription regulator RofA related
1	gbs1530	Rga transcription regulator RofA related
14	gbs1582	transcriptional regulator (LysR family)
20	gbs1671	two-component sensor histidine kinase
15	gbs1736	transcriptional regulator (LacI family)
13	gbs1807	transcriptional regulator, MerR/GlnR family
5	gbs1834	DltS two-component sensor histidine kinase
5	gbs1835	DltR two-component response regulator
20	gbs1842	transcriptional regulator (antiterminator)
27	gbs1849	transcriptional regulator (LacI family)
20	gbs1943	two-component sensor histidine kinase (C-terminal part)
20	gbs1944	two-component response regulator
27	gbs1972	transcriptional regulator (phage related)
29	gbs2055	arginine repressor ArgR
27	gbs2117	transcriptional regulator
27	gbs2120	transcriptional regulator (Crp/Fnr family)

## **II. Partie sur l'étude d'une nouvelle famille d'ICEs les TnGBS.**

**Publication n°6**

---

## Genome Sequence of *Streptococcus gallolyticus*: Insights into Its Adaptation to the Bovine Rumen and Its Ability To Cause Endocarditis<sup>∇†</sup>

Christophe Rusniok,<sup>2</sup> Elisabeth Couvé,<sup>1</sup> Violette Da Cunha,<sup>1</sup> Rachida El Gana,<sup>1#</sup>  
Nora Zidane,<sup>3</sup> Christiane Bouchier,<sup>3</sup> Claire Poyart,<sup>5</sup> Roland Leclercq,<sup>6</sup>  
Patrick Trieu-Cuot,<sup>4</sup> and Philippe Glaser<sup>1\*</sup>

Institut Pasteur, Laboratoire Evolution et Génomique Bactérienne, CNRS URA 2171,<sup>1</sup> Institut Pasteur, Unité de Biologie des Bactéries Intracellulaires, CNRS URA 2171,<sup>2</sup> and Institut Pasteur, Genopole Plate-Forme Génomique,<sup>3</sup> and Institut Pasteur, Unité de Biologie des Bactéries Pathogènes à Gram Positif, CNRS URA 2172,<sup>4</sup> 28 rue du Dr Roux 75724 Paris, Centre National de Référence des Streptocoques, APHP, INSERM 567, Université Paris Descartes, 75014 Paris,<sup>5</sup> and Service de Microbiologie and EA 2128 Relations Hôte et Microorganismes des Épithéliums, Hôpital Côte de Nacre, Université de Caen, 14033 Caen,<sup>6</sup> France

Received 20 December 2009/Accepted 1 February 2010

*Streptococcus gallolyticus* (formerly known as *Streptococcus bovis* biotype I) is an increasing cause of endocarditis among streptococci and frequently associated with colon cancer. *S. gallolyticus* is part of the rumen flora but also a cause of disease in ruminants as well as in birds. Here we report the complete nucleotide sequence of strain UCN34, responsible for endocarditis in a patient also suffering from colon cancer. Analysis of the 2,239 proteins encoded by its 2,350-kb-long genome revealed unique features among streptococci, probably related to its adaptation to the rumen environment and its capacity to cause endocarditis. *S. gallolyticus* has the capacity to use a broad range of carbohydrates of plant origin, in particular to degrade polysaccharides derived from the plant cell wall. Its genome encodes a large repertoire of transporters and catalytic activities, like tannase, phenolic compounds decarboxylase, and bile salt hydrolase, that should contribute to the detoxification of the gut environment. Furthermore, *S. gallolyticus* synthesizes all 20 amino acids and more vitamins than any other sequenced *Streptococcus* species. Many of the genes encoding these specific functions were likely acquired by lateral gene transfer from other bacterial species present in the rumen. The surface properties of strain UCN34 may also contribute to its virulence. A polysaccharide capsule might be implicated in resistance to innate immunity defenses, and glucan mucopolysaccharides, three types of pili, and collagen binding proteins may play a role in adhesion to tissues in the course of endocarditis.

Several studies have reported that the proportion of infective endocarditis due to *Streptococcus gallolyticus* has increased during the last decades, concomitantly with a decrease of cases due to oral streptococci (35). *S. gallolyticus* is now becoming the first cause of infectious endocarditis among streptococci in Europe (16). Furthermore, *S. gallolyticus* endocarditis is associated with rural residency, suggesting transmission from animals (29). However, the reasons for the emergence of this pathogen remain poorly understood. *S. gallolyticus* belongs to the *Streptococcus bovis* group known for more than 60 years to cause endocarditis (45). Recently, the former species *S. bovis* has been divided into four major species (50, 53). *S. gallolyticus* corresponds to *S. bovis* biotype I (mannitol fermentation positive), the closely related species *S. pasteurianus* to biotype II/2 (mannitol negative and  $\beta$ -glucuronidase positive), and the

more distantly related species *S. infantarius* to biotype II/1 (mannitol negative and  $\beta$ -glucuronidase negative). *S. macedonicus*, the fourth species, commonly found in cheese, is non-pathogenic and also considered a *S. gallolyticus* subspecies (53, 62). A majority of endocarditis cases was due, among the formerly *S. bovis* group, to *S. gallolyticus* strains (4).

Multiple studies have shown that endocarditis due to *S. gallolyticus* as well as positive blood culture for this species is often associated with gastrointestinal malignancy (4, 6). This association has led to a strong indication for gastrointestinal investigation and endoscopic follow-up in the case of *S. gallolyticus* infections (66). The association of *S. gallolyticus* infection with colon cancer is a major but still unsolved issue. It may be just incidental, as the alteration of the digestive mucosa may favor the translocation of the bacteria into the bloodstream. Alternatively, the tumor may contribute to the proliferation of *S. gallolyticus* in close proximity to the gut epithelium, increasing its probability of translocating through the gut barrier. It has also been suggested that the bacterium itself contributes to carcinogenesis (60, 69). In addition to human disease, *S. gallolyticus* may also cause diseases in animals, like septicemia in pigeons (19), outbreaks in broiler flocks (11), or bovine mastitis (28).

Independent from its association to disease, *S. gallolyticus*

\* Corresponding author. Mailing address: Laboratoire Evolution et Génomique Bactérienne, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris cedex 15, France. Phone: 33 1 45 68 89 96. Fax: 33 1 45 68 87 86. E-mail: pglaser@pasteur.fr.

# Present address: Biorad Laboratories, 3 boulevard Raymond Poincaré, 92430 Marnes la Coquette, France.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

<sup>∇</sup> Published ahead of print on 5 February 2010.



has been isolated as a tannin-resistant bacterium from the feces of different mammalian herbivores, including the koala (48) or the Japanese large wood mouse (52), and it is also a normal inhabitant of the rumen (39). Its resistance to tannins is linked to its tannase activity, a characteristic which also led this bacterium to be named “*gallyoticus*” as it is able to decarboxylate gallate, an organic acid derived from tannin degradation. *S. gallyoticus* is also known to express other degradative functions unique among streptococci, like a bile salt hydrolase or an amylase. These properties allow its multiplication outside the animal host, as *S. gallyoticus* was isolated from a digester fed with shea cake (derived from the nuts of the African tree *Vitellaria paradoxa*) rich in tannins and aromatic compounds (12). *S. gallyoticus* is a commensal of the human intestinal tract but remains a rarely detected (2.5 to 15%) low-abundance species (10, 40). In herbivores, overgrowth of *S. bovis* may become deleterious. For example, ingestion of large amounts of rapidly fermented cereal grains leads to a destabilization of the rumen flora and to the proliferation of acid-tolerant bacteria, including *S. gallyoticus*. This is accompanied by the overproduction of mucopolysaccharides that stabilize the foam, resulting in feedlot bloat, a significant cause of economical loss (14).

Virulence and colonization factors of *S. gallyoticus* in humans are largely unknown. Studies of the bird host have shown that this *Streptococcus* species expresses a capsular polysaccharide, and five different serotypes have been described (19). In addition, electron microscopy studies have revealed the presence of fimbria-like structures on the surface of *S. gallyoticus*. It was hypothesized that capsules and/or fimbriae are involved in virulence (63). *S. gallyoticus* isolates responsible for endocarditis exhibited heterogeneous patterns of adherence to extracellular matrix (ECM) proteins, which suggests that they produce different surface components (55). Recently, a collagen binding adhesin together with 10 putative ECM binding proteins were identified in the draft genome sequence of a human isolate of *S. gallyoticus* (54).

Here we describe the sequence and analysis of the genome of *S. gallyoticus* strain UCN34 isolated from a human case of endocarditis associated with colon cancer. Analysis of the predicted proteins revealed unique metabolic and cell surface features among streptococci, which contribute to its adaptation to the rumen and to its ability to cause endocarditis. We showed by comparative genomics that many of the corresponding genes were probably acquired by lateral gene transfer (LGT) from other *Firmicutes* of the gut microbiota.

#### MATERIALS AND METHODS

**Bacterial strains, plasmids, and growth conditions.** *S. gallyoticus* subsp. *gallyoticus* strain UCN34 was isolated in 2001 at the Hospital in Caen (Calvados, France) and is resistant to tetracycline. It was recovered from blood cultures of a 70-year-old man with a 1-month history of intermittent fever. The patient was found to have endocarditis of his native aortic valve. He was successfully treated with amoxicillin and gentamicin. Subsequent digestive endoscopy revealed a colic cancer, which led to a partial colectomy. For the construction of the large insert library and the shotgun libraries, pSYX34 and XL10-blue Kan<sup>r</sup> (Stratagene) or pcDNA2.1 (Invitrogen) and XL2-blue were used as vectors or recipient strains, respectively. *Escherichia coli* and *S. gallyoticus* strains were grown in Luria and brain heart infusion (BHI) broth, respectively.

**Sequencing and assembly methods.** Genome sequencing was performed using the conventional whole-genome shotgun strategy (24, 26). Two libraries (1- to 2-kb and 2- to 3-kb inserts) were generated by random mechanical shearing of

genomic DNA and cloning into pcDNA-2.1 (Invitrogen). A scaffold was obtained by end sequencing clones from a medium-size insert library (5 to 10 kb) in the low-copy-number vector pSYX34 (68). Recombinant plasmids were used as templates for cycle sequencing reactions consisting of 35 cycles (96°C for 30 s; 50°C for 15 s; 60°C for 4 min) in a thermocycler. Samples were precipitated and loaded onto 96-lane automatic capillary 3700 and 3730 DNA sequencers (Applied Biosystems). In an initial step, 30,240 sequences from the three libraries were assembled into 144 contigs using the Phred/Phrap/Consed software (22, 32). CAAT Box (25) was used to predict links between contigs. PCR amplification products amplified from UCN34 chromosomal DNA as template were used to fill gaps and to resequence low-quality regions using primers designed by Consed. If no link was predicted between contigs, direct sequencing with chromosomal DNA as template was performed. Sequence reactions consisted of 99 cycles (96°C for 30 s; 50°C for 15 s; 60°C for 4 min) in a thermocycler. A total of 250 mM betaine was added to the reaction mixture.

**Annotation methods.** The CAAT Box environment (25) was used for genome annotation. Coding sequences (CDS) were defined by combining GeneMark predictions (38) with visual inspection of each open reading frame (ORF) for the presence of a start codon with an upstream ribosome binding site and BlastP similarity searches of the Nrprot database (2). The GeneMark predictions were trained on a set of ORFs longer than 300 codons encoding proteins similar to proteins with known function present in public databases. Initially, only CDS longer than 80 codons were retained. Subsequently, all CDS between 40 and 80 codons were searched using the same matrix, but only those with a high coding probability, as predicted by GeneMark, were retained. In a final step, all intergenic regions were searched for short or truncated genes by BlastX comparisons with protein sequence libraries. Limits of the rRNA operons were identified by homology with the other streptococcal genomes; tRNAs were searched using tRNAscan-SE (41).

All predicted CDS were examined visually. Function predictions were based on BlastP similarity searches and on the analysis of motifs using the PFAM databases. Toppred2 was used to identify transmembrane domains (15), SignalP version 2.0 to predict signal peptide regions (44), and LipPred to predict lipoproteins (58). Orthologs between *S. gallyoticus* UCN34 and *Streptococcus agalactiae* (strain NEM316) (31), *Streptococcus suis* (strain 05ZYH33) (13), *Streptococcus thermophilus* (strain CNRZ1066) (7), *Streptococcus mutans* (strain UA159) (1), *Streptococcus sanguinis* (strain SK36) (67), *Streptococcus uberis* (strain 0140J) (64), *Streptococcus pyogenes* (strain M1) (23), *Streptococcus pneumoniae* (strain TIGR4) (59), and *Streptococcus zooepidemicus* (strain MGCS10565) (5) were defined as genes showing bidirectional best hits by *S. gallyoticus* proteome BlastP comparisons. The threshold was set to a minimum of 50% sequence identity and a ratio of 0.8 to 1.25 of the protein length. Multiple sequence alignments were done using T-Coffee, and the phylogenetic trees by a Bayesian analysis using MrBayes at the phylogeny website (<http://www.phylogeny.fr/>) (20).

**Nucleotide sequence accession number.** The *S. gallyoticus* UCN34 genome sequence is available from DDBJ/GenBank/EMBL under accession number FN597254.

#### RESULTS AND DISCUSSION

**General properties and comparative genomics.** The genome of *S. gallyoticus* strain UCN34 consists of a single circular chromosome of 2,350,911 bp (Fig. 1). It is second in size among streptococcal strains for which complete genome sequences are publicly available, being only 37 kb smaller than the *S. sanguinis* SK36 genome (67). The G+C content of the genome is 37.6%. It contains six rRNA operons, 71 tRNA genes, and 2,239 protein coding genes, 13 of which are pseudogenes. The average gene size is 306 codons, accounting for 87% of the genome. These values are similar to the average values reported for the genus *Streptococcus* (5). We assigned functional annotations to 1,586 proteins, whereas 574 proteins were annotated as conserved hypothetical proteins. *S. gallyoticus* UCN34 encodes only 79 proteins without similarity to proteins described in the sequence databases (E value threshold of e-3). This small number of orphan genes probably reflects both the large number of sequenced streptococcal genomes and the relative paucity of UCN34 in mobile genetic elements, which commonly carry orphan genes (17). However, 454 genes, i.e.,

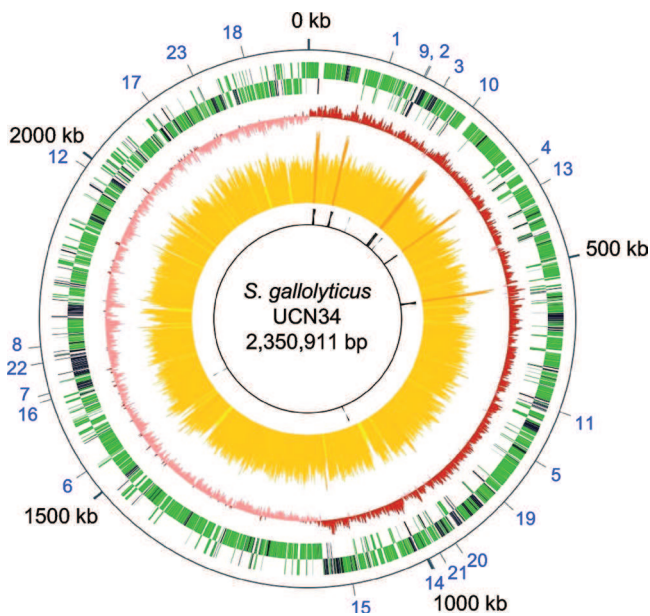


FIG. 1. Circular genome map of *S. gallolyticus* UCN34 showing the position and orientation of predicted genes. From outside to inside, circle 1, protein coding genes on the + and - strands, with genes with nonstreptococcal best BlastP hit in black; circle 2, G/C bias (G+C/G-C); circle 3: G+C content, with <30.9% G+C in yellow, between 30.9% and 44.4% G+C in orange, and with >44.4% G+C in red; circle 4, stable RNA coding genes. Numbers in blue on the external circle correspond to genes discussed in the text: polysaccharide utilization (1, *gallo\_0112*; 2, *gallo\_0162*; 3, *gallo\_0189*; 4, *gallo\_0330*; 5, *gallo\_0757*; 6, *gallo\_1462*; 7, *gallo\_1577-1578*; 8, *amyE*), vitamin biosynthesis (9, *panBCD*; 10, *panE*; 11, *ribDEAH*; 12, *bioBDY*), polysaccharide biosynthesis (13, *gallo\_0364-0367*; 14, *cps* operon; 15, *gallo\_1052-1057*), pilus operons (16, *gallo\_1568-1570*; 17, *gallo\_2040-2038*; 18, *gallo\_2179-2177*), and detoxification (19, *bsh*; 20, *gallo\_0906*; 21, *tanA*; 22, *gallo\_1609*; 23, *gallo\_2106*). The scale in kb is indicated outside the genome, with the predicted origin of replication being at position 0.

one-fifth of the predicted genes, encode proteins showing a best BlastP hit with proteins from nonstreptococcal species. Most of them are clustered in genomic regions of up to 61 genes (in black in Fig. 1). The best BlastP hits were predominantly with other *Firmicutes* species, principally with enterococci, lactobacilli, bacilli, and clostridia, bacterial species also present in the rumen microbiota and in the human gut (Fig. 2). This result suggests that these genes were acquired by LGT. Analysis of the functional categories of these genes revealed that they are enriched in functions related to transport, carbohydrate metabolism, cofactor biosynthesis, detoxification, and regulation, which may account for the specific adaptation of *S. gallolyticus* to the rumen environment (see Fig. S1 in the supplemental material).

To get a better insight into the relationship of *S. gallolyticus* with other streptococcal species, we compared its genome sequence with those of 10 other completely sequenced streptococcal species (Table 1). *S. gallolyticus* showed the largest number of orthologous genes with *S. agalactiae*, a gut-adapted pyogenic *Streptococcus* species and a major cause of neonatal infections that also causes mastitis in the bovine host, and with *S. mutans*, a *Streptococcus* species of the viridans group responsible for dental caries. A high proportion of best BlastP hits

was also found with *S. salivarius*, a commensal *Streptococcus* species from the dorsum of the tongue and the saliva belonging also to the viridans group (Fig. 2). Therefore, *S. gallolyticus* shares properties mainly with two oral streptococci from the viridans group and with the gut-associated species *S. agalactiae*. As these different species do not cluster phylogenetically (57), this gene content conservation likely reflects conserved lifestyle and niche adaptation rather than phylogenetic proximity. In contrast, 687 genes of strain UCN34 did not show any ortholog among the 10 other sequenced streptococci. Among these genes, 244 have, nevertheless, a best BlastP hit with a streptococcal gene. This observation indicates that they have been acquired by LGT from other streptococci or that they are paralogous genes. In addition to having transport, carbohydrate metabolism, and regulatory functions, they are enriched in genes coding for surface proteins and proteins implicated in cell wall biosynthesis (see Fig. S2 in the supplemental material). These functions are probably involved in the interaction of *S. gallolyticus* with the gut environment and its ability to cause disease in humans (see below).

***S. gallolyticus* metabolizes a broad range of carbohydrates of plant origin.** The most remarkable feature of the catabolism of *S. gallolyticus* as deduced from the genome sequence is that it should have the capacity to degrade diverse complex polysaccharides (Table 2). Strain UCN34 was predicted to utilize plant storage carbohydrates, like various forms of alpha glycoside polymers, including starch and glycogen. Furthermore the genome encodes nine members of the maltogenic alpha amylase family (glycosidase family 13), including a secreted alpha amylase homologous to the *Bacillus subtilis amyE* gene product. In addition, *S. gallolyticus* UCN34 should be able to utilize polymers of fructose, levan, and inulin, which are storage polysaccharides in plants, as we identified a gene encoding a secreted membrane-bound fructan hydrolase (*fruA*). Unique among streptococci is the presence of genes encoding secreted enzymes predicted to be involved in the degradation of insoluble polysaccharides of the plant cell wall. The gene *gallo\_0162* encodes a putative secreted mannanase with a C-terminal domain similar to known *Bacillus* mannanases, whereas the N-terminal domain is a carbohydrate-binding domain, and gene *gallo\_0330* encodes a secreted protein highly similar to *Ruminococcus albus* endoglucanase V with cellulase activity (47). Strain UCN34 also synthesizes two putative secreted pectate lyases (*Gallo\_1577* and *Gallo\_1578*). The catalytic domain of these two paralogous proteins is similar to that of the pectate lyase from *Clostridium acetobutylicum*. Finally, *gallo\_0189* encodes a putative secreted arabinogalactan endo-1,4-beta-galactosidase similar to a protein from *Enterococcus faecium*. These genes encoding glycosyl hydrolases are located in chromosomal regions sharing few orthologs with other streptococci, suggesting that they might have been gained by LGT.

It is known that *S. gallolyticus* is able to use a broad range of carbohydrates, like cellobiose, fructose, galactose, glucose, lactose, maltose, mannitol, melibiose, raffinose, and trehalose (12). In order to investigate the predictions made from the genome analysis, we further tested the capacity of *S. gallolyticus* to utilize different carbon sources using the API-50 assay (bioMérieux). This showed that strain UCN34 is indeed able to ferment mannose, *N*-acetylglucosamine, salicin, sucrose, inulin, starch, glycogen, and gentobiose (data not shown). In

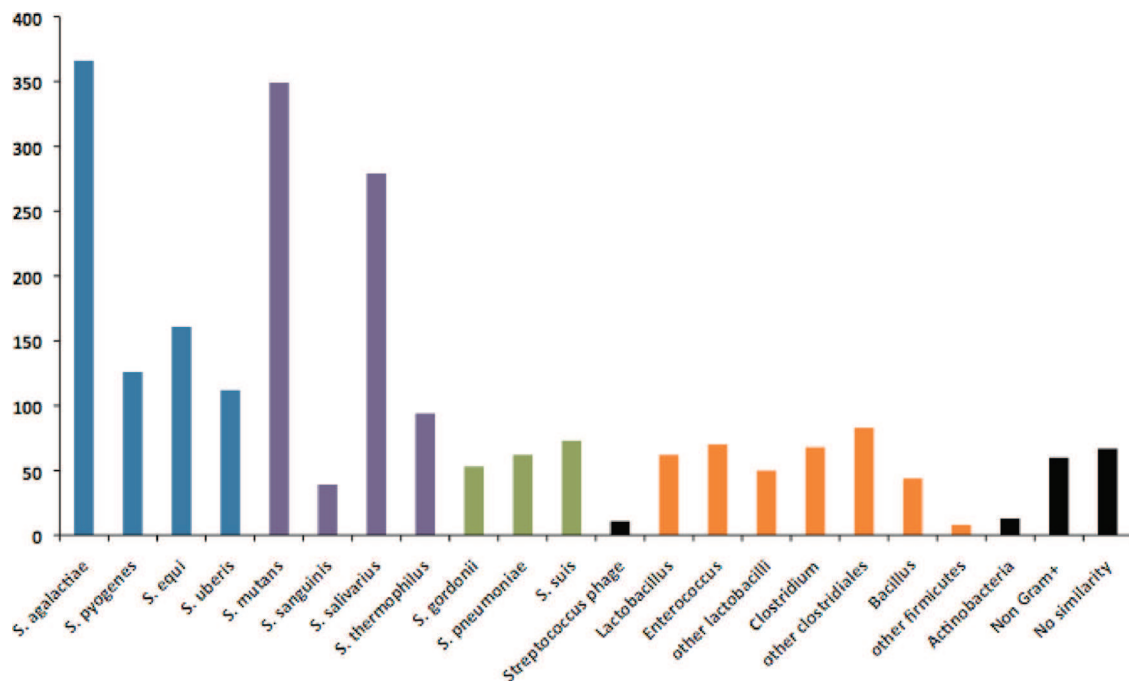


FIG. 2. Distribution of best BlastP hits. Blue, pyogenic streptococci; purple, viridans group streptococci; green, other streptococci; orange, other *Firmicutes*; black, others.

agreement with these observations, genome analysis revealed an extremely broad range of sugar permeases, as among the 322 genes possibly involved in transport, 80 were predicted as specific for carbon sources. As an example, strain UCN34 expresses 25 transporters of the phosphotransferase systems (PTS), whereas *S. mutans* and *S. uberis*, two species known for their ability to ferment numerous sugars, express only 14 and 15 PTS transporters, respectively (1, 64). These 25 PTS transporters belong to the seven categories defined by Barabote and Saier according to their specificity for different sugars (3): glucose (10), lactose (5), ascorbate (3), fructose (2), mannose (2), galactitol (2), and glucitol (1). Strain UCN34 encodes a great diversity of enzymes to metabolize these carbohydrates, including 33 hydrolases belonging to nine different families. In particular, it encodes 11 phospho-beta-glucosidases (family 1) involved in degrading plant-derived polysaccharides, which is a number similar to that reported for *S. uberis*, which is known to utilize metabolites deriving from the degradation of plant poly-

saccharides (64). In addition to the ability to utilize and degrade plant sugars, *S. gallolyticus* is adapted to use other nutrients of plant origin, as it is the only streptococcus able to utilize malate, a keto acid that occupies a central role in plant metabolism. Strain UCN34 expresses a malolactic enzyme (encoded by *gallo\_2048*) associated with a malate transporter (encoded by *gallo\_2049*).

Taken together, the specific catabolic capacities of *S. gallolyticus* identified from its genome sequence suggest that they provide a selective advantage to this bacterium for life in the gut environment of herbivores, as it contains a large diversity of compounds of plant origin. *S. gallolyticus* should be able to degrade them and does not depend on other microorganisms for the initial degradation of insoluble plant polysaccharides which are abundant in this environment.

***S. gallolyticus* has only few nutritional requirements.** Although most streptococci are auxotrophic for several amino

TABLE 1. Genome comparison

Parameter	<i>S. agalactiae</i> NEM316	<i>S. pyogenes</i> SF370	<i>S.</i> <i>zooepidemicus</i> MGCS10565	<i>S. uberis</i> 0140	<i>S. mutans</i> UA159	<i>S. sanguinis</i> SK36	<i>S. gordonii</i> Challis	<i>S. pneumoniae</i> TIGR4	<i>S. thermophilus</i> CNRZ1066	<i>S. suis</i> 05ZYH33
No. of orthologs	1,128	974	1,008	1,095	1,211	1,052	1,051	972	1,048	954
No. of CDS	2,094	1,697	1,893	1,825	1,960	2,270	2,051	2,104	1,915	2,186
% of orthologs <sup>a</sup>	54	57	53	60	62	46	51	46	55	44
No. of best orthologs	372	115	124	134	386	40	59	47	215	56
No. of orthologs identified in a single genome	40	15	13	18	44	6	7	12	25	24
No. of best BlastP hits <sup>b</sup>	392	130	163	119	382	45	57	67	258	79

<sup>a</sup> According to the target genome.

<sup>b</sup> In this species, by BlastP search against the NCBI nr protein database.

TABLE 2. Secreted polysaccharide hydrolases

Gene name	Predicted activity	Best blast hit	Binding domain
Plant cell wall constituents			
<i>gallo_0162</i>	Mannanase	<i>Bacillus</i>	Lectin domain
<i>gallo_1578</i>	Pectate lyase	<i>Clostridium acetobutylicum</i>	
<i>gallo_1577</i>	Pectate lyase	<i>Clostridium acetobutylicum</i>	Lectin domain
<i>gallo_0189</i>	Endo-beta-1,4-galactanase	<i>Enterococcus faecium</i>	
<i>gallo_0330</i>	Beta-1,4-endoglucanase V (cellulase)	<i>Ruminococcus albus</i>	
Plant reserve polysaccharides			
<i>gallo_0757</i>	Alpha-amylase	<i>S. equi</i>	Starch binding
<i>gallo_1462</i>	Pullulanase	<i>S. suis</i>	LPXTG motif
<i>amyE</i>	Alpha-amylase	<i>B. subtilis</i>	
<i>gallo_0112</i>	Fructan hydrolase	<i>S. sanguinis</i>	LPXTG motif

acids and vitamins, previous experiments have shown that most *S. bovis* strains show an absolute requirement only for biotin, while thiamine stimulates growth, and no requirement for any amino acid (45). In agreement with this observation, we identified from the genome of strain UCN34 the complete biosynthetic pathways for the 20 amino acids (aa) and for selenocysteine. Comparison of the distribution of these genes among streptococci (Table 3) showed that streptococci isolated from the oral cavity, such as *S. mutans*, *S. sanguinis*, *S. gordonii*, and *S. pneumoniae*, have the capacity to synthesize most amino acids. However, *S. gallolyticus* and *S. mutans* are the only two *Streptococcus* species expressing the glutamate synthase genes (*gltAB*) clustered in an operon with the glutamine synthetase gene (*glnA*). The glutamate synthase is a key enzyme in conjunction with glutamine synthetase for the assimilation of ammonium ions into cellular metabolic pathways. The genetic organizations of the amino acid biosynthetic pathways are

identical in *S. gallolyticus* and in *S. mutans*, which is also a prototroph for all amino acids. The analysis of the distribution of the different amino acid biosynthetic pathways among the 11 *Streptococcus* species shown in Table 3 suggests that the common ancestor of the genus was proficient in synthesizing all amino acids, and then during evolution, a heterogeneous gene loss took place in the different *Streptococcus* species, leading to different synthesis capacities. Among nonoral streptococci, *S. gallolyticus* is particular, as it has kept the ability to synthesize all 20 amino acids.

In agreement with the growth requirements, we identified the complete pathways for the synthesis of multiple vitamins, including riboflavin (B2), nicotine amide (NAD, B3), pantothenate (B5), pyridoxine (B6), and folic acid (B9), and partial biosynthetic pathways for biotin (B8) and thiamine (B1) (Table 4). *S. gallolyticus* is the only *Streptococcus* species that possesses complete biosynthetic pathways for pantothenate and NAD,

TABLE 3. Amino acid biosynthetic pathways in 11 *Streptococcus* species

<i>S. gallolyticus</i> UCN34	<i>S. agalactiae</i> NEM316	<i>S. pyogenes</i> M1GAS	<i>S. zooepidemicus</i> MGCS10565	<i>S. uberis</i> 0140	<i>S. mutans</i> UA159	<i>S. sanguinis</i> SK36	<i>S. gordonii</i> Challis	<i>S. pneumoniae</i> TIGR4	<i>S. thermophilus</i> CNRZ1066	<i>S. suis</i> 05ZYH33
<i>argCJBD</i>	<i>argF</i>	<i>argF</i>	<i>argF</i>	<i>argF</i>	<i>argCJBD</i>	<i>argCJBD</i>	<i>argCJBD</i>	<i>argF</i>	<i>argCJBD</i>	<i>argF</i>
<i>argF</i>	<i>argGH</i>			<i>argGH</i>	<i>argF</i>	<i>argF</i>	<i>argF</i>	<i>argF</i>	<i>argF</i>	<i>argGH</i>
<i>argGH</i>					<i>argGH</i>	<i>argGH</i>	<i>argGH</i>	<i>argGH</i>	<i>argGH</i>	
<i>aroFBDC</i>	<i>aroFBC</i>	<i>aroFC</i>	<i>aroFC</i>	<i>aroFC</i>	<i>aroFBDC</i>	<i>aroFBDC</i>	<i>aroFBDC</i>	<i>aroFBDC</i>	<i>aroFBDC</i>	<i>aroFBDC</i>
<i>aroEK</i>	<i>aroEK</i>	<i>aroEK</i>	<i>aroEK</i>	<i>aroEK</i>	<i>aroEK</i>	<i>aroEK</i>	<i>aroEK</i>	<i>aroEK</i>	<i>aroEK</i>	<i>aroK</i>
<i>pheA</i>					<i>pheA</i>	<i>pheA</i>	<i>pheA</i>	<i>pheA</i>	<i>pheA</i>	
<i>trpEGDCFBA</i>	<i>trpG</i>	<i>trpG</i>	<i>trpG</i>		<i>trpEGDCFBA</i>	<i>trpEGDCFBA</i>	<i>trpEGDCFBA</i>	<i>trpEGDCFBA</i>	<i>trpEGDCFBA</i>	
<i>tyrA</i>					<i>tyrA</i>	<i>tyrA</i>	<i>tyrA</i>	<i>tyrA</i>	<i>tyrA</i>	<i>tyrA</i>
<i>cysEK</i>	<i>cysEK</i>	<i>cysEK</i>	<i>cysEK</i>	<i>cysEK</i>	<i>cysEK</i>	<i>cysEK</i>	<i>cysEK</i>	<i>cysEK</i>	<i>cysEK</i>	<i>cysEK</i>
<i>dapA</i>				<i>dapA</i>	<i>dapA</i>	<i>dapA</i>	<i>dapA</i>	<i>dapA</i>	<i>dapA</i>	<i>dapA</i>
<i>dapB</i>				<i>dapB</i>	<i>dapB</i>	<i>dapB</i>	<i>dapB</i>	<i>dapB</i>	<i>dapB</i>	<i>dapB</i>
<i>dapD</i>				<i>dapD</i>	<i>dapD</i>	<i>dapD</i>	<i>dapD</i>	<i>dapD</i>	<i>dapD</i>	<i>dapD</i>
<i>lysA</i>				<i>lysA</i>	<i>lysA</i>	<i>lysA</i>	<i>lysA</i>	<i>lysA</i>	<i>lysA</i>	<i>lysA</i>
<i>glnA</i>	<i>glnA</i>	<i>glnA</i>	<i>glnA</i>	<i>glnA</i>	<i>glnA</i>	<i>glnA</i>	<i>glnA</i>	<i>glnA</i>	<i>glnA</i>	<i>glnA</i>
<i>gltAB</i>					<i>gltAB</i>					
<i>glyA</i>	<i>glyA</i>	<i>glyA</i>	<i>glyA</i>	<i>glyA</i>	<i>glyA</i>	<i>glyA</i>	<i>glyA</i>	<i>glyA</i>	<i>glyA</i>	<i>glyA</i>
<i>hisCGDBHAFI</i>					<i>hisCGDBHAFI</i>	<i>hisCGDBHAFI</i>	<i>hisCGDBHAFI</i>			
<i>ilvDBHCA</i>					<i>ilvDBHCA</i>	<i>ilvDBHCA</i>	<i>ilvDBHCA</i>	<i>ilvDBHCA</i>	<i>ilvDBHCA</i>	<i>ilvDBHCA</i>
<i>leuABCD</i>					<i>leuABCD</i>	<i>leuABCD</i>	<i>leuABCD</i>	<i>leuA</i>	<i>leuABCD</i>	<i>leuABCD</i>
<i>metA</i>	<i>metE</i>	<i>metK</i>	<i>metK</i>	<i>metK</i>	<i>metA</i>	<i>metA</i>	<i>metA</i>	<i>metA</i>	<i>metA</i>	<i>metA</i>
<i>metC</i>	<i>metK</i>				<i>metC</i>	<i>metC</i>	<i>metC</i>	<i>metC</i>	<i>metC</i>	<i>metC</i>
<i>metE</i>					<i>metE</i>	<i>metK</i>	<i>metK</i>	<i>metK</i>	<i>metK</i>	<i>metK</i>
<i>metK</i>					<i>metK</i>					
<i>proBA</i>	<i>proBA</i>	<i>proBA</i>	<i>proBA</i>	<i>proBA</i>	<i>proBA</i>	<i>proA</i>	<i>proA</i>	<i>proA</i>	<i>proBA</i>	<i>proA</i>
<i>proI</i>	<i>proI</i>	<i>proI</i>	<i>proI</i>	<i>proI</i>	<i>proI</i>				<i>proI</i>	
<i>serCA</i>	<i>serCA</i>			<i>serCA</i>	<i>serCA</i>	<i>serCA</i>	<i>serB</i>		<i>serCA</i>	<i>serCA</i>
<i>serB</i>	<i>serB</i>			<i>serB</i>	<i>serB</i>	<i>serB</i>			<i>serB</i>	<i>serB</i>
<i>hom</i>	<i>hom thrBC</i>		<i>hom</i>	<i>hom thrBC</i>	<i>hom</i>	<i>hom</i>	<i>hom</i>	<i>hom</i>	<i>hom</i>	<i>hom thrBC</i>
<i>thrBC</i>			<i>thrBC</i>		<i>thrBC</i>	<i>thrBC</i>	<i>thrBC</i>	<i>thrBC</i>	<i>thrBC</i>	

TABLE 4. Vitamin biosynthetic pathways in 11 *Streptococcus* species

<i>S. gallolyticus</i> UCN34	<i>S. agalactiae</i> NEM316	<i>S. pyogenes</i> MIGAS	<i>S.</i> <i>zooepidemicus</i> MGCS10565	<i>S. uberis</i> 0140	<i>S. mutans</i> UA159	<i>S. sanguinis</i> SK36	<i>S. gordonii</i> Challis	<i>S. pneumoniae</i> TIGR4	<i>S. thermophilus</i> CNRZ1066	<i>S. suis</i> 05ZYH33
<i>BioBDY</i>	<i>bioB</i>									
<i>folCEPBK</i>	<i>folCEPBK</i>	<i>folCEPBK</i>	<i>folCEPBK</i>	<i>folCEPBK</i>	<i>folCEPBK</i>	<i>folCEP</i>	<i>folE</i>	<i>folE</i>	<i>folCEPBK</i>	<i>folCEPBK</i>
<i>folD</i>	<i>folD</i>	<i>folD</i>	<i>folD</i>	<i>folD</i>	<i>folD</i>	<i>folD</i>	<i>folD</i>	<i>folD</i>	<i>folD</i>	<i>folD</i>
<i>nadABC</i>										
<i>nadE</i>	<i>nadE</i>	<i>nadE</i>	<i>nadE</i>	<i>nadE</i>	<i>nadE</i>	<i>nadE</i>	<i>nadE</i>	<i>nadE</i>	<i>nadE</i>	<i>nadE</i>
<i>panBCD</i>									<i>panE</i>	
<i>panE</i>										
<i>ribDEAH</i>	<i>ribDEAH</i>							<i>ribDEAH</i>		
<i>ribF</i>	<i>ribF</i>	<i>ribF</i>	<i>ribF</i>	<i>ribF</i>	<i>ribF</i>	<i>ribF</i>	<i>ribF</i>	<i>ribF</i>	<i>ribF</i>	<i>ribF</i>
<i>tenA thiEMD</i>	<i>tenA thiEMD</i>							<i>thiEMD</i>		<i>thiEMD</i>
<i>thiI</i>	<i>thiI</i>	<i>thiI</i>	<i>thiI</i>	<i>thiI</i>	<i>thiI</i>	<i>thiI</i>	<i>thiI</i>	<i>thiI</i>	<i>thiI</i>	<i>thiI</i>
<i>thiN</i>	<i>thiN</i>	<i>thiN</i>	<i>thiN</i>	<i>thiN</i>	<i>thiN</i>	<i>thiN</i>	<i>thiN</i>	<i>thiN</i>	<i>thiN</i>	<i>thiN</i>

whereas riboflavin biosynthesis is present only in *S. gallolyticus*, *S. agalactiae*, and *S. pneumoniae*. Pantothenate biosynthesis involves four enzymatic steps from  $\alpha$ -ketoisovalerate and L-aspartate catalyzed by the products of *panB*, *-C*, *-D*, and *-E*. The *panBCD* operon of *S. gallolyticus* is highly similar to those present in various clostridial species, bacteria that are also present in the rumen. In contrast, the *panE* gene encoding a 2-dehydropantoate 2-reductase is located elsewhere on the chromosome and has an ortholog only in *S. thermophilus*. The high identity of the two *panE* genes (97% at the nucleotide level) indicates a recent event of LGT between the two streptococcal species. *S. gallolyticus* is also able to synthesize NAD from aspartate in three steps. The *nadABC* operon, encoding a quinolinate synthetase, an L-aspartate oxidase, and a nicotinate-nucleotide pyrophosphorylase, respectively, is organized again similarly to counterparts found in various clostridial species present in the rumen. Finally, phylogenetic analysis of the *ribDEAH* operon, encoding the enzymes catalyzing the four steps of riboflavin biosynthesis in *S. gallolyticus*, *S. agalactiae*, and *S. pneumoniae* revealed that they do not cluster on the evolutionary tree and that they were thus probably gained independently by LGT (see Fig. S3 in the supplemental material). These genes, like the gene involved in pantothenate and NAD biosynthesis, are clustered in regions predicted to have been acquired by LGT. Although we cannot rule out that all other species have lost these genes, we suggest that unlike with the genes coding amino acid biosynthesis pathways, the above-described functions were acquired from other gut bacteria by LGT and the common ancestor was likely auxotrophic for these vitamins.

The genome analysis suggests that due to the very few nutritional requirements of *S. gallolyticus* it can grow in an environment containing a diverse range of carbohydrates and poor in amino acids and vitamins, probably enabling it to outcompete auxotrophic bacteria in the rumen and in the human colon.

***S. gallolyticus* expresses diverse and unique functions among streptococci that are necessary to survive in hostile environments.** *S. gallolyticus* has a particularly versatile lifestyle, with the capacity to adapt to different environments and to survive harsh conditions. It colonizes different mammalian and bird hosts, causes a broad range of diseases, and was also isolated outside a host as the major species in a continuous digester fed

with shea cake (12). In agreement with its adaptation capacities, the genome of strain UCN34 encodes more regulatory proteins than any other streptococcal genome sequenced to date, as 177 genes (7.7% of all predicted genes) are devoted to regulatory functions. This proportion is slightly higher than that of *L. monocytogenes* (7.3%) (30) and close to that found in *Pseudomonas aeruginosa* (8.4%) (56), which are both environmental opportunistic pathogens characterized by their capacity to colonize diverse environments.

Another specific characteristic of *S. gallolyticus* that is important in hostile environments is its ability to degrade tannins, which are toxic polyphenolic compounds that form strong complexes with proteins and other macromolecules. Indeed, a gene, *tanA*, encoding a 596-aa-long protein 43% identical to the tannase recently described for *Staphylococcus lugdunensis* (46), is present in its genome. Like *S. lugdunensis* tannase, TanA is predicted to be exported and contains a conserved lipobox motif (LTACS) at the cleavage site of the signal peptide, indicating that it may be a lipoprotein that remains attached to the cell membrane. Strain UCN34 also expresses a nonsecreted protein similar to TanA (Gallo\_1609). This protein has homologs in diverse *Firmicutes* species and may have a similar hydrolase activity. *tanA* and *gallo\_1609* have no counterpart in other streptococcal genomes and were probably gained by LGT.

Degradation of hydrolyzable tannins by *S. gallolyticus* produces a phenolic compound, gallic acid, which may also be toxic. However, *S. gallolyticus* is able to decarboxylate gallate and to use it as an alternative carbon supply (12). This capacity may be achieved by two decarboxylases encoded in the genome that have no orthologs in other sequenced streptococcal genomes. Gallo\_2106 (PadC) is predicted to be a phenolic acid decarboxylase similar to those present in lactobacilli and bacilli. For example, the *B. subtilis* PadC protein decarboxylates phenolic compounds and is implicated in the phenolic acid stress response (61). Gallo\_0906 belongs to the carboxymuconolactone decarboxylase family. Both activities may account for the ability of *S. gallolyticus* to decarboxylate gallate but also other phenolic compounds, like protocatechuic *p*-coumaric or caffeic and ferulic acids, as previously described (12).

Another important feature enabling *S. gallolyticus* to multiply in the gut environment is its capacity to hydrolyze bile salt conferring resistance to this detergent. This capacity is proba-

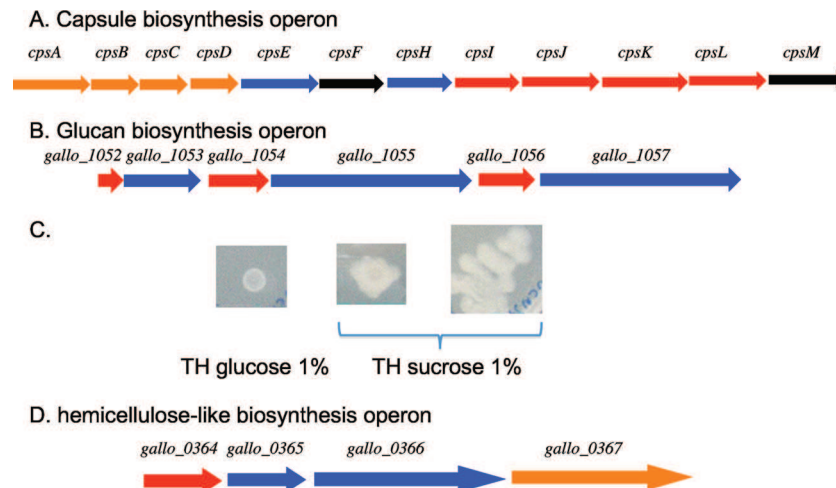


FIG. 3. Loci involved in the synthesis of extracellular polysaccharides. (A) The 12-gene operon responsible for the synthesis of the capsular polysaccharides. *cpsA* (*gallo\_0944*) encodes a regulator of the LytR family, *cpsE* (*gallo\_0948*) an undecaprenyl-phosphate glycosyl-1-phosphate transferase, *cpsF* (*gallo\_0949*) a rhamosyl transferase, *cpsH* (*gallo\_0950*) a polymerase, *cpsI* and *cpsJ* (*gallo\_0951*) sugar transferases, *cpsK* (*gallo\_0953*) the transporter or flippase, *cpsL* (*gallo\_0954*) a CDP-glycerol:polyglycerol phosphate glycerol-phosphotransferase, and *cpsM* (*gallo\_0955*) a glycerol-3-phosphate cytidyltransferase. The color of the arrow indicates the species encoding the best BlastP hit: red, *S. pneumoniae*; blue, *S. agalactiae*; orange, *S. thermophilus*; and black, other. (B) Putative locus involved in mucopolysaccharide biosynthesis. Genes encoding glycosyltransferases are in blue and regulators in red. (C) *S. gallolyticus* UCN34 grown on TH medium supplemented with 1% glucose or sucrose. (D) Operon putatively involved in the biosynthesis of a hemicellulose-like molecule. *gallo\_0364* encodes a putative diguanylate cyclase, *gallo\_0365* and *gallo\_0366* encode two putative glycosyl transferases, and *gallo\_0367* encodes a 12-transmembrane domain protein.

bly linked to the *bsh* (*gallo\_0818*) gene encoding a protein highly similar to *Listeria monocytogenes* bile salt hydrolase (63% identity) (21). This activity, not yet described for another streptococcus, is commonly found in diverse bacteria dominant in the gut community, like *Clostridia*, *Lactobacillus*, or *Bacteroidetes* organisms.

In addition to these enzymatic activities, *S. gallolyticus* expresses a broad range of transport systems possibly involved in the adaptation to diverse environments. We predicted 25 genes encoding efflux proteins likely involved in detoxification. Strain UCN34 expresses six efflux proteins belonging to the multidrug and toxic compound extrusion (MATE) family, whereas other streptococci (*S. gordonii* and *S. uberis*) encode not more than three efflux proteins of this family. Furthermore, strain UCN34 encodes 58 proteins probably transporting inorganic compounds. Among those, eight have no counterpart in other sequenced streptococcal genomes (including a sulfate family and two heavy-metal transporters). It also expresses three different iron transport systems; *gallo\_0590* encodes an NRMAMPn<sup>++</sup>, Fe<sup>++</sup> transporter; *gallo\_1771-1774* an iron ABC transporter; and *gallo\_0619-0620* a ferrous iron transporter. Strain UCN34 is, among the sequenced streptococci, the only one to express all three systems, conferring to *S. gallolyticus* probably a higher capacity to capture iron, a compound rare in the digestive tract and the environment.

In the gut environment, bacteria need to defend themselves against diverse viruses and other mobile genetic elements (MGE) that may aggress them. CRISPR small RNAs are important in the bacterial response to these invaders. Strikingly, strain UCN34 carries two CRISPR loci located next to each other on the chromosome that belongs to the two major classes of CRISPR loci (8). The first one, with 16 spacer sequences, is associated with three *cas* (CRISPR-associated) genes,

*gallo\_1439* to *gallo\_1437*. The second locus contains 14 spacers and four *cas* genes, *gallo\_1443* to *gallo\_1446*. Together with *S. thermophilus*, strain UCN34 is the only *Streptococcus* species whose genome sequence is known that encodes multiple CRISPR systems. They may contribute to its resistance to phages and other mobile elements.

***S. gallolyticus* synthesizes diverse polysaccharides.** *S. gallolyticus* is known to produce an extracellular capsule considered a virulence factor. We identified an operon containing 12 genes encoding proteins similar to enzymes involved in the biosynthesis of capsular polysaccharides of different streptococci, in particular of *S. pneumoniae* and *S. thermophilus* exopolysaccharides (Fig. 3A). Its organization and encoded proteins showed a high degree of similarity with those of the *S. pneumoniae* serotype 23F capsule, although the best BlastP hits for the polymerase (encoded by *cpsH*) and for the undecaprenyl-phosphate glycosyl-1-phosphate transferase (encoded by *cpsE*) were to the serotype VIII capsular proteins of *S. agalactiae*. Thus, this operon probably encodes the functions for the synthesis of a capsular structure of *S. gallolyticus*, which might confer its resistance to complement and to innate immunity, allowing survival in blood as described for these two pathogenic streptococci (37, 43). It has also been shown that *S. gallolyticus* expresses the human sialyl Lewis antigen, which may contribute to the ability of this bacterium to cross the vascular endothelium (34). The above-described locus may encode the polysaccharide mimicking the human sialyl Lewis antigen.

It was shown that the production of mucopolysaccharides by *S. bovis* increases the viscosity of ruminal fluid and stabilizes the foam implicated in frothy feedlot bloat (14). A second locus identified in the UCN34 genome encodes three different glycosyl transferases where each structural gene is associated

with a regulatory gene (Fig. 3B). *gallo\_1053*, encoding a non-secreted transferase, is preceded by an *abrB*-like regulatory gene, and the two paralogous genes *gallo\_1055* and *gallo\_1057*, encoding secreted transferases, highly similar to the three glycosyltransferases, GtfA, GtfB, and GtfC, of *S. mutans*, are each preceded by an *rgg*-like regulatory gene. In *S. mutans*, the three extracellular glycosyltransferases are involved in the biosynthesis of insoluble glucans from sucrose. These glucans mediate the adherence of the bacterial cells to the tooth surface and contribute to biofilm formation. The similarity of the genetic organization and the clustering of the three glycosyl transferase genes in *S. gallolyticus* strongly support the hypothesis that these genes were acquired by LGT from oral streptococci. In order to learn whether strain UCN34 is able to produce glucans, we compared growth on Todd-Hewitt (TH) medium supplemented either with glucose or sucrose (Fig. 3C). Whereas the colonies on TH-glucose medium were small and nonshiny, they were extremely mucoid on medium supplemented with sucrose. This suggests that the glucans produced by these glycosyl transferases represent the mucopolysaccharides produced during feedlot bloat in cattle (14). Thus, among streptococci, *S. gallolyticus* is the only one known to produce both a capsule and a mucoid extracellular glucan.

Finally, we also identified an operon (*gallo\_0364-0367*) possibly involved in the biosynthesis of hemicellulose (Fig. 3D). These four genes encode two glycosyltransferases, including a putative bacterial cellulose synthase (Gallo\_0366), a transmembrane (TM) protein (Gallo\_0367), and a putative diguanylate cyclase (Gallo\_0364). Gallo\_0364 contains the characteristic GGDEF motif and carries, in addition, a PAS domain, indicating that it is probably a sensory protein. Diguanylate cyclases are shown to have important regulatory roles in many bacteria. However, Gallo\_0364 is the first example of such a protein described for a streptococcal species. Furthermore, cellulose biosynthesis in Gram-negative bacteria has been shown to depend on cyclic di-GMP (51) and is involved in biofilm formation (27). The polysaccharide produced by this biosynthetic pathway may be part of the external matrix of a *S. gallolyticus* biofilm.

#### Strain UCN 34 expresses diverse surface proteins and pili.

Together with the diversity of surface-exposed polysaccharides, a large repertoire of genes encoding surface proteins belonging to different families is present in the *S. gallolyticus* genome. Many of them may be involved in its colonization capacity and virulence. For example, strain UCN34 encodes a putative fibrinogen/fibronectin binding protein (Fbp) and four proteins related to staphylococcal collagen binding proteins (Gallo\_0577, Gallo\_1570, Gallo\_2032, and Gallo\_2179). However, among these four proteins, only Gallo\_2179 carries the collagen binding motif and likely binds collagen. Collagen binding has been shown to be required for the capacity of *Staphylococcus aureus* to cause endocarditis (33). Similarly, these proteins may contribute to the capacity of strain UCN34 to colonize the endocardium.

Nineteen proteins possess both an N-terminal signal peptide and a C-terminal LPXTG sorting motif. Among these proteins, three have predicted enzymatic functions: a subtilisin-like serine protease (Gallo\_0748), a pullulanase (Gallo\_1462), and a fructan hydrolase (FruA). Furthermore, three operons encoding a sortase C and two LPXTG motif proteins (*gallo\_1570-*

*1568*, *gallo\_2040-2038*, and *gallo\_2179-2177*) were identified. These gene clusters present several characteristics of pilus operons, indicating that strain UCN34 may have the capacity to synthesize three different types of pilus appendages. Unlike Gram-negative bacteria, Gram-positive bacteria polymerize pilus subunits by transpeptidation reactions catalyzed by sortase C. These pili of Gram-positive bacteria have been shown to contribute to the colonization of specific host tissues, the modulation of host immune responses, and the development of bacterial biofilms (42). In *S. gallolyticus* UCN34, two of these pilus operons express a protein similar to collagen binding proteins (Gallo\_1570 and Gallo\_2179), predicted to be adhesins. Similarly, it was recently described for *S. gallolyticus* strain TX20005 that it encodes three pilus operons and also collagen binding proteins (54). The locus *gallo\_2179-2177* is identical to the *acb* locus of strain TX20005, and the locus *gallo\_2040-2038* is identical to the *sbs15* locus. In contrast, the third locus (*gallo\_1570-1568*) is only distantly related to the third locus of strain TX20005 (*sbs13*). This indicates diversity in the surface protein repertoire among *S. gallolyticus* isolates.

Expression of multiple sortase C proteins associated to pilus loci is common in streptococci. However, strain UCN34 presents the unique property of expressing three highly similar sortase A-encoding genes. In addition to the housekeeping *srtA* gene located downstream the *gyrA* gene, we identified two paralogous genes both located on mobile genetic elements. Gallo\_1651 is carried by a predicted integrative and conjugative element (ICE) (ICESgal1; encoded by *gallo\_1646* to *gallo\_1703*) that carries also a Tn916-like conjugative transposon (encoded by *gallo\_1676-int* to *gallo\_1700*) expressing the *tet(M)* determinant. Furthermore, within Tn916, the previously described plasmid pBC16 which carried the *tet(L)* determinant is inserted (49). This presence of *tet(M)* and *tet(L)* explains the resistance of strain UCN34 to tetracycline. Interestingly, ICESgal1 expresses two LPXTG proteins (Gallo\_1649 and Gallo\_1675), which may contribute to the conjugative transfer of the element. The third sortase gene, *gallo\_0299*, is also located on an MGE, TnGallo1. TnGallo1 carries also three genes encoding LPXTG proteins and is similar to the recently described conjugative transposon TnGBS2 of *S. agalactiae* (9). Phylogenetic analysis of the amino acid sequences of these three sortases indicates that they are closely related and clustered within the housekeeping SrtA sequences of streptococci. Therefore, the two additional sortase genes were probably acquired by LGT of *srtA* genes from related species (Fig. 4). It is tempting to suggest that these additional sortase A genes contribute to the propagation of MGEs by anchoring surface proteins encoded by these elements and involved in their conjugative transfer; however, they may also contribute to the cell wall anchoring of proteins encoded by the genome backbone and to the general fitness of the cell.

Another important class of surface proteins is lipoproteins. The genome of strain UCN34 encodes 42 lipoproteins (see Table S1 in the supplemental material). Twenty-six are ABC transporter substrate binding proteins, one is the above-described tannase, one is a peptidyl-prolyl *cis-trans* isomerase, and 14 are conserved hypothetical proteins. Further analysis of the deduced protein sequences revealed in 27 out of the 42 lipoproteins the presence of a serine-rich motif following the lipid-modified cysteine residue (Table S1). For example, the

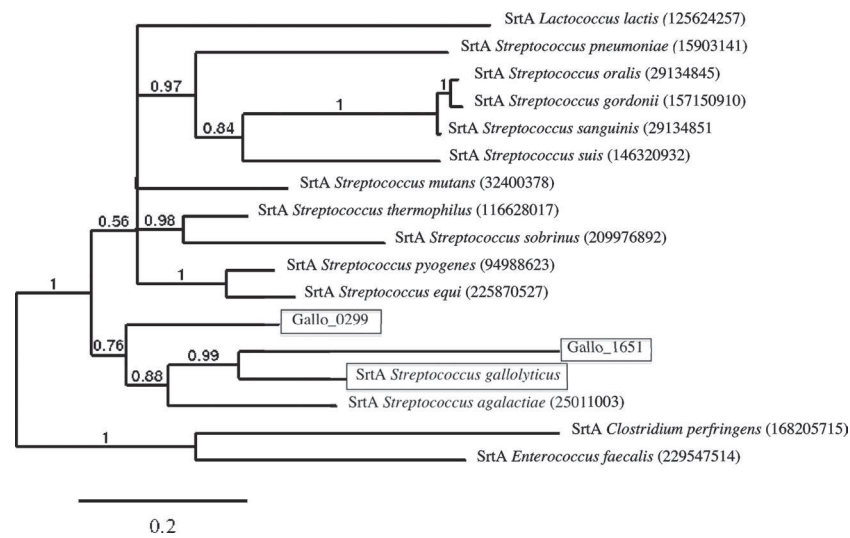


FIG. 4. Phylogenetic relationship of sortase A protein sequences with representative sequences from *Firmicutes*. The three sortase A encoded by strain UCN34 are boxed. Gallo\_0299 is expressed by TnGallo1 and Gallo\_1651 by ICESgal1. SrtA is the *S. gallolyticus* housekeeping sortase encoded by the *srtA* gene located downstream of the *gyrA* gene. The scale bar represents 0.2 substitutions per site. The numbers at the branches are posterior probabilities indicating the support for the branch. GI numbers for sequences are given in parentheses.

cysteine residue in Gallo\_1845 is followed by 11 serine residues. Interestingly, alignments with similar proteins from other streptococci revealed that the serine-rich domain is in most cases not conserved. This indicated that this motif might be specific for *S. gallolyticus* lipoproteins and recently acquired. We thus searched for such motifs in the *S. gallolyticus* proteins. We identified 16 proteins containing one or several serine-rich domains that were not conserved in orthologous proteins of other streptococci. These serine-rich regions are systematically found in domains predicted to be extracellular (Table S1). Similar polyserine repeat regions have been identified with plant cell wall-degrading enzymes of environmental bacteria, like *Cellvibrio japonicus* (18), *Microbulbifer degradans* (36), or *Saccharophagus degradans* (65). It has been proposed that these flexible spacer regions enhance substrate accessibility for its degradation. Interestingly, in *S. gallolyticus*, these polyserine repeats are found in surface-exposed proteins with diverse predicted functions not directly related to carbohydrate degradation. Some of them are involved in cell wall biosynthesis (penicillin binding proteins 1A and 2C, D-alanyl-D-alanine-carboxypeptidase, and polyglycerol synthase) or in regulation (protein kinase). They are thus possibly linked to specific interactions of the bacteria with polysaccharides from the environment or that they produce.

**Conclusion.** *S. gallolyticus* is an important cause of endocarditis; still very little is known about the genetic basis of virulence and niche adaptation. Analysis of the *S. gallolyticus* genome sequence has revealed important features that might help to understand its virulence and survival strategies. We identified an impressive number of functions that are specific to the species *S. gallolyticus* among streptococci but that are shared with *Lactobacillus*, *Bacillus*, and clostridial species belonging to the normal rumen flora, suggesting that *S. gallolyticus* acquired different functions by LGT from other inhabitants of the rumen to better adapt to this environment. These functions may also provide some explanation for the association of

*S. gallolyticus* and colon cancer. In a healthy human gut, the low prevalence of *S. gallolyticus* compared to what is observed with ruminants is probably linked to the difference in diet and to the different environments in the rumen and the human gut. However, alterations of the gut due to colon cancer or neoplastic polyps may affect the flux of intestinal content, leading to the accumulation of material, including plant-derived fibers in close proximity to the epithelial cells and to the tumor. This material is metabolically poor but rich in fiber carbohydrates and possibly tannins. Thus, this may now represent a favorable microenvironment for the proliferation of *S. gallolyticus*. Our genome analysis therefore substantiates the model of overgrowth of *S. gallolyticus* linked to colon dysplasia and also supports the observation that the prevalence of *S. bovis* in fecal cultures from patients with carcinoma of the colon was significantly increased compared to that in controls (10, 40). Furthermore, this proliferation together with the alteration of the gut epithelium may favor translocation of the bacteria to the bloodstream. Subsequently, the predicted capsular polysaccharides will protect *S. gallolyticus* from innate immunity responses, and the surface proteins and pili showing ECM binding properties, including collagen binding, will aid its adherence to endothelium cells as a preliminary step in the development of infective endocarditis.

The genome sequence of strain UCN34 reported here, isolated from human blood, is the first obtained for a *S. gallolyticus* isolate. It will provide the genomic information for designing a multilocus sequence typing scheme to study the population of *S. gallolyticus* strains in more detail. This will lead to a better definition of the relationship between strains isolated from humans, bovines, and birds. This will also be the basis for a rational strain selection for further genome sequencing using high-throughput methods to identify putative genomics and functional specificities associated with the different hosts.



## ACKNOWLEDGMENTS

This work was supported by the Institut Pasteur Genopole program. V.D. is supported by a grant from the Ministère Délégué à l'Enseignement Supérieur et à la Recherche (France).

We wish also to thank Mathieu Brochet, Carmen Buchrieser, Maria-Jose Lopez, Harold Tjalsma, and Isabelle Rosinski-Chupin for critical comments.

## REFERENCES

- Ajdic, D., W. M. McShan, R. E. McLaughlin, G. Savic, J. Chang, M. B. Carson, C. Primeaux, R. Tian, S. Kenton, H. Jia, S. Lin, Y. Qian, S. Li, H. Zhu, F. Najjar, H. Lai, J. White, B. A. Roe, and J. J. Ferretti. 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *Proc. Natl. Acad. Sci. U. S. A.* **99**:14434–14439.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Barabote, R. D., and M. H. Saier, Jr. 2005. Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiol. Mol. Biol. Rev.* **69**:608–634.
- Beck, M., R. Frodl, and G. Funke. 2008. Comprehensive study of strains previously designated *Streptococcus bovis* consecutively isolated from human blood cultures and emended description of *Streptococcus gallolyticus* and *Streptococcus infantarius* subsp. *coli*. *J. Clin. Microbiol.* **46**:2966–2972.
- Beres, S. B., R. Sesso, S. W. Pinto, N. P. Hoe, S. F. Porcella, F. R. Deleo, and J. M. Musser. 2008. Genome sequence of a Lancefield group C *Streptococcus zooepidemicus* strain causing epidemic nephritis: new information about an old disease. *PLoS One* **3**:e3026.
- Boleij, A., R. M. Schaeps, and H. Tjalsma. 2009. Association between *Streptococcus bovis* and colon cancer. *J. Clin. Microbiol.* **47**:516.
- Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau, and P. Hols. 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat. Biotechnol.* **22**:1554–1558.
- Bolotin, A., B. Quinquis, A. Sorokin, and S. D. Ehrlich. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**:2551–2561.
- Brochet, M., V. Da Cunha, E. Couve, C. Rusniok, P. Trieu-Cuot, and P. Glaser. 2009. Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. *Mol. Microbiol.* **71**:948–959.
- Burns, C. A., R. McCaughey, and C. B. Lauter. 1985. The association of *Streptococcus bovis* fecal carriage and colon neoplasia: possible relationship with polyps and their premalignant potential. *Am. J. Gastroenterol.* **80**:42–46.
- Chadfield, M. S., J. P. Christensen, A. Decostere, H. Christensen, and M. Bisgaard. 2007. Geno- and phenotypic diversity of avian isolates of *Streptococcus gallolyticus* subsp. *gallolyticus* (*Streptococcus bovis*) and associated diagnostic problems. *J. Clin. Microbiol.* **45**:822–827.
- Chamkha, M., B. K. Patel, A. Traore, J. L. Garcia, and M. Labat. 2002. Isolation from a shea cake digester of a tannin-degrading *Streptococcus gallolyticus* strain that decarboxylates protocatechuic and hydroxycinnamic acids, and emendation of the species. *Int. J. Syst. Evol. Microbiol.* **52**:939–944.
- Chen, C., J. Tang, W. Dong, C. Wang, Y. Feng, J. Wang, F. Zheng, X. Pan, D. Liu, M. Li, Y. Song, X. Zhu, H. Sun, T. Feng, Z. Guo, A. Ju, J. Ge, Y. Dong, W. Sun, Y. Jiang, J. Yan, H. Yang, X. Wang, G. F. Gao, R. Yang, and J. Yu. 2007. A glimpse of streptococcal toxic shock syndrome from comparative genomics of *S. suis* 2 Chinese isolates. *PLoS One* **2**:e315.
- Cheng, K. J., T. A. McAllister, J. D. Popp, A. N. Hristov, Z. Mir, and H. T. Shin. 1998. A review of bloat in feedlot cattle. *J. Anim. Sci.* **76**:299–308.
- Claros, M. G., and G. von Heijne. 1994. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10**:685–686.
- Corredoira, J., M. P. Alonso, A. Coira, E. Casariego, C. Arias, D. Alonso, J. Pita, A. Rodriguez, M. J. Lopez, and J. Varela. 2008. Characteristics of *Streptococcus bovis* endocarditis and its differences with *Streptococcus viridans* endocarditis. *Eur. J. Clin. Microbiol. Infect. Dis.* **27**:285–291.
- Daubin, V., and H. Ochman. 2004. Start-up entities in the origin of new genes. *Curr. Opin. Genet. Dev.* **14**:616–619.
- DeBoy, R. T., E. F. Mongodin, D. E. Fouts, L. E. Tailford, H. Khouri, J. B. Emerson, Y. Mohamoud, K. Watkins, B. Henrissat, H. J. Gilbert, and K. E. Nelson. 2008. Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. *J. Bacteriol.* **190**:5455–5463.
- De Herdt, P., F. Haesebrouck, L. A. Devriese, and R. Ducatelle. 1992. Biochemical and antigenic properties of *Streptococcus bovis* isolated from pigeons. *J. Clin. Microbiol.* **30**:2432–2434.
- Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie, and O. Gascuel. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**:W465–469.
- Dussurget, O., D. Cabanes, P. Dehoux, M. Lecuit, C. Buchrieser, P. Glaser, and P. Cossart. 2002. *Listeria monocytogenes* bile salt hydrolase is a PrfA-regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. *Mol. Microbiol.* **45**:1095–1106.
- Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**:175–185.
- Ferretti, J. J., W. M. McShan, D. Ajdic, D. J. Savic, G. Savic, K. Lyon, C. Primeaux, S. Sezate, A. N. Suvorov, S. Kenton, H. S. Lai, S. P. Lin, Y. Qian, H. G. Jia, F. Z. Najjar, Q. Ren, H. Zhu, L. Song, J. White, X. Yuan, S. W. Clifton, B. A. Roe, and R. McLaughlin. 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl. Acad. Sci. U. S. A.* **98**:4658–4663.
- Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**:496–512.
- Franguel, L., P. Glaser, C. Rusniok, C. Buchrieser, E. Duchaud, P. Dehoux, and F. Kunst. 2004. CAAT-Box, Contigs-Assembly and Annotation Tool-Box for genome sequencing projects. *Bioinformatics* **20**:790–797.
- Franguel, L., K. E. Nelson, C. Buchrieser, A. Danchin, P. Glaser, and F. Kunst. 1999. Cloning and assembly strategies in microbial genome projects. *Microbiology* **145**:2625–2634.
- Garcia, B., C. Latasa, C. Solano, F. Garcia-del Portillo, C. Gamazo, and I. Lasa. 2004. Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. *Mol. Microbiol.* **54**:264–277.
- Garvie, E. I., and A. J. Bramley. 1979. *Streptococcus bovis*—an approach to its classification and its importance as a cause of bovine mastitis. *J. Appl. Bacteriol.* **46**:557–566.
- Giannitsioti, E., C. Chirouze, A. Bouvet, I. Beguinot, F. Delahaye, J. L. Mainardi, M. Celard, L. Mihaila-Amrouche, V. L. Moing, and B. Hoen. 2007. Characteristics and regional variations of group D streptococcal endocarditis in France. *Clin. Microbiol. Infect.* **13**:770–776.
- Glaser, P., L. Franguel, C. Buchrieser, C. Rusniok, A. Amend, F. Baquero, P. Berche, H. Bloeker, P. Brandt, T. Chakraborty, A. Charbit, F. Chetouani, E. Couve, A. de Daruvar, P. Dehoux, E. Domann, G. Dominguez-Bernal, E. Duchaud, L. Durant, O. Dussurget, K. D. Entian, H. Fsihi, F. G. Portillo, P. Garrido, L. Gautier, W. Goebel, N. Gomez-Lopez, T. Hain, J. Hauf, D. Jackson, L. M. Jones, U. Kaerst, J. Kreff, M. Kuhn, F. Kunst, G. Kurapkat, E. Madueno, A. Maitournam, J. M. Vicente, E. Ng, H. Nedjari, G. Nordtsiek, S. Novella, B. de Pablos, J. C. Perez-Diaz, R. Purcell, B. Rimmel, M. Rose, T. Schlueter, N. Simoes, A. Tierrez, J. A. Vazquez-Boland, H. Voss, J. Wehland, and P. Cossart. 2001. Comparative genomics of *Listeria* species. *Science* **294**:849–852.
- Glaser, P., C. Rusniok, C. Buchrieser, F. Chevalier, L. Franguel, 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.
- Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**:195–202.
- Hienz, S. A., T. Schennings, A. Heimdahl, and J. I. Flock. 1996. Collagen binding of *Staphylococcus aureus* is a virulence factor in experimental endocarditis. *J. Infectious Dis.* **174**:83–88.
- Hirota, K., R. Osawa, K. Nemoto, T. Ono, and Y. Miyake. 1996. Highly expressed human sialyl Lewis antigen on cell surface of *Streptococcus gallolyticus*. *Lancet* **347**:760.
- Hoen, B., C. Chirouze, C. H. Cabell, C. Selton-Suty, F. Duchene, L. Olaison, J. M. Miro, G. Habib, E. Abrutyn, S. Eykyn, Y. Bernard, F. Marco, and G. R. Corey. 2005. Emergence of endocarditis due to group D streptococci: findings derived from the merged database of the International Collaboration on Endocarditis. *Eur. J. Clin. Microbiol. Infect. Dis.* **24**:12–16.
- Howard, M. B., N. A. Ekborg, L. E. Taylor, S. W. Hutcheson, and R. M. Weiner. 2004. Identification and analysis of polyserine linker domains in prokaryotic proteins with emphasis on the marine bacterium *Microbulbifer degradans*. *Protein Sci.* **13**:1422–1425.
- Hyams, C., E. Camberlein, J. M. Cohen, K. Bax, and J. S. Brown. 2010. The *Streptococcus pneumoniae* capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect. Immun.* **78**:704–715.
- Isono, K., J. D. McIninch, and M. Borodovsky. 1994. Characteristic features of the nucleotide sequences of yeast mitochondrial ribosomal protein genes as analyzed by computer program GeneMark. *DNA Res.* **1**:263–269.
- Khafipour, E., S. Li, J. C. Plaizier, and D. O. Krause. 2009. Rumen microbiome composition determined using two nutritional models of subacute ruminal acidosis. *Appl. Environ. Microbiol.* **75**:7115–7124.
- Klein, R. S., R. A. Recco, M. T. Catalano, S. C. Edberg, J. I. Casey, and N. H. Steigbigel. 1977. Association of *Streptococcus bovis* with carcinoma of the colon. *New Engl. J. Med.* **297**:800–802.
- Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved

- detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
42. Mandlik, A., A. Swierczynski, A. Das, and H. Ton-That. 2008. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol.* **16**:33–40.
  43. Marques, M. B., D. L. Kasper, M. K. Pangburn, and M. R. Wessels. 1992. Prevention of C3 deposition by capsular polysaccharide is a virulence mechanism of type III group B streptococci. *Infect. Immun.* **60**:3986–3993.
  44. Nielsen, H., S. Brunak, and G. von Heijne. 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* **12**:3–9.
  45. Niven, C. F., M. R. Washburn, and J. C. White. 1948. Nutrition of *Streptococcus bovis*. *J. Bacteriol.* **55**:601–606.
  46. Noguchi, N., T. Ohashi, T. Shiratori, K. Narui, T. Hagiwara, M. Ko, K. Watanabe, T. Miyahara, S. Taira, F. Moriyasu, and M. Sasatsu. 2007. Association of tannase-producing *Staphylococcus lugdunensis* with colon cancer and characterization of a novel tannase gene. *J. Gastroenterol.* **42**:346–351.
  47. Ohara, H., J. Noguchi, S. Karita, T. Kimura, K. Sakka, and K. Ohmiya. 2000. Sequence of egV and properties of EgV, a *Ruminococcus albus* endoglucanase containing a dockerin domain. *Biosci. Biotechnol. Biochem.* **64**:80–88.
  48. Osawa, R. 1990. Formation of a clear zone on tannin-treated brain heart infusion agar by a *Streptococcus* sp. isolated from feces of koalas. *Appl. Environ. Microbiol.* **56**:829–831.
  49. Palva, A., G. Vigren, M. Simonen, H. Rintala, and P. Laamanen. 1990. Nucleotide sequence of the tetracycline resistance gene of pBC16 from *Bacillus cereus*. *Nucleic Acids Res.* **18**:1635.
  50. Poyart, C., G. Quesne, and P. Trieu-Cuot. 2002. Taxonomic dissection of the *Streptococcus bovis* group by analysis of manganese-dependent superoxide dismutase gene (sodA) sequences: reclassification of '*Streptococcus infantarius subsp. coli*' as *Streptococcus lutetiensis* sp. nov. and of *Streptococcus bovis* biotype 11.2 as *Streptococcus pasteurianus* sp. nov. *Int. J. Syst. Evol. Microbiol.* **52**:1247–1255.
  51. Romling, U. 2002. Molecular biology of cellulose production in bacteria. *Res. Microbiol.* **153**:205–212.
  52. Sasaki, E., T. Shimada, R. Osawa, Y. Nishitani, S. Spring, and E. Lang. 2005. Isolation of tannin-degrading bacteria isolated from feces of the Japanese large wood mouse, *Apodemus speciosus*, feeding on tannin-rich acorns. *Syst. Appl. Microbiol.* **28**:358–365.
  53. Schlegel, L., F. Grimont, P. A. Grimont, and A. Bouvet. 2004. New group D streptococcal species. *Indian J. Med. Res.* **119**(Suppl.):252–256.
  54. Sillanpää, J., S. R. Nallapareddy, X. Qin, K. V. Singh, D. M. Muzny, C. L. Kovar, L. V. Nazareth, R. A. Gibbs, M. J. Ferraro, J. M. Steckelberg, G. M. Weinstock, and B. E. Murray. 2009. A collagen-binding adhesin, Acb, and 10 other putative MSCRAMM and pilus family proteins of *Streptococcus galloyticus* subsp. *galloyticus* (*S. bovis* biotype I). *J. Bacteriol.* **191**:6643–6653.
  55. Sillanpää, J., S. R. Nallapareddy, K. V. Singh, M. J. Ferraro, and B. E. Murray. 2008. Adherence characteristics of endocarditis-derived *Streptococcus galloyticus* ssp. *galloyticus* (*Streptococcus bovis* biotype I) isolates to host extracellular matrix proteins. *FEMS Microbiol. Lett.* **289**:104–109.
  56. Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**:959–964.
  57. Tapp, J., M. Thollessen, and B. Herrmann. 2003. Phylogenetic relationships and genotyping of the genus *Streptococcus* by sequence determination of the RNase P RNA gene, *mipB*. *Int. J. Syst. Evol. Microbiol.* **53**:1861–1871.
  58. Taylor, P. D., C. P. Toseland, T. K. Attwood, and D. R. Flower. 2006. LIPPRED: a web server for accurate prediction of lipoprotein signal sequences and cleavage sites. *Bioinformatics* **1**:176–179.
  59. Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**:498–506.
  60. Tjalsma, H., M. Scholler-Guinard, E. Lasonder, T. J. Ruers, H. L. Willems, and D. W. Swinkels. 2006. Profiling the humoral immune response in colon cancer patients: diagnostic antigens from *Streptococcus bovis*. *Int. J. Cancer* **119**:2127–2135.
  61. Tran, N. P., J. Gury, V. Dartois, T. K. Nguyen, H. Seraut, L. Barthelmebs, P. Gervais, and J. F. Cavin. 2008. Phenolic acid-mediated regulation of the *padC* gene, encoding the phenolic acid decarboxylase of *Bacillus subtilis*. *J. Bacteriol.* **190**:3213–3224.
  62. Tsakalidou, E., E. Zoidou, B. Pot, L. Wassill, W. Ludwig, L. A. Devriese, G. Kalantzopoulos, K. H. Schleifer, and K. Kersters. 1998. Identification of streptococci from Greek Kasseri cheese and description of *Streptococcus macedonicus* sp. nov. *Int. J. Syst. Bacteriol.* **48**:519–527.
  63. Vanrobaeys, M., P. De Herdt, G. Charlier, R. Ducatelle, and F. Haesebrouck. 1999. Ultrastructure of surface components of *Streptococcus galloyticus* (*S. bovis*) strains of differing virulence isolated from pigeons. *Microbiology* **145**:335–342.
  64. 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, and J. Parkhill. 2009. Evidence for niche adaptation in the genome of the bovine pathogen *Streptococcus uberis*. *BMC Genomics* **10**:54.
  65. Weiner, R. M., L. E. Taylor II, B. Henrissat, L. Hauser, M. Land, P. M. Coutinho, C. Rancurel, E. H. Saunders, A. G. Longmire, H. Zhang, E. A. Bayer, H. J. Gilbert, F. Larimer, I. B. Zhulin, N. A. Ekborg, R. Lamed, P. M. Richardson, I. Borovok, and S. Hutcheson. 2008. Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2-40 T. *PLoS Genet.* **4**:e1000087.
  66. Wentling, G. K., P. P. Metzger, E. J. Dozois, H. K. Chua, and M. Krishna. 2006. Unusual bacterial infections and colorectal carcinoma—*Streptococcus bovis* and *Clostridium septicum*: report of three cases. *Dis. Colon Rectum* **49**:1223–1227.
  67. Xu, P., J. M. Alves, T. Kitten, A. Brown, Z. Chen, L. S. Ozaki, P. Manque, X. Ge, M. G. Serrano, D. Puiu, S. Hendricks, Y. Wang, M. D. Chaplin, D. Akan, S. Paik, D. L. Peterson, F. L. Macrina, and G. A. Buck. 2007. Genome of the opportunistic pathogen *Streptococcus sanguinis*. *J. Bacteriol.* **189**:3166–3175.
  68. Xu, S. Y., and A. Fomenkov. 1994. Construction of pSC101 derivatives with Cam<sup>r</sup> and Tet<sup>r</sup> for selection or LacZ<sup>'</sup> for blue/white screening. *Biotechniques* **17**:57.
  69. zur Hausen, H. 2006. *Streptococcus bovis*: causal or incidental involvement in cancer of the colon? *Int. J. Cancer* **119**:xi–xii.

## RESUME

*Streptococcus agalactiae*, également appelé Streptocoque du Groupe B ou GBS, est la première cause de septicémies et de méningites chez les nouveaux-nés. Cette bactérie à Gram positif est également responsable d'infections chez les personnes âgées et les personnes immunodéprimées. On la retrouve aussi associée à des infections chez l'animal. Néanmoins, c'est avant tout une bactérie commensale du tube digestif présente chez 20 % de la population humaine. Mon travail de thèse porte sur deux stratégies utilisées par *S. agalactiae* pour s'adapter à la variabilité de son environnement: la mise en place de réseaux de régulation associés à la perception de cet environnement et la modification de son patrimoine génétique notamment grâce aux transferts horizontaux faisant intervenir des éléments génétiques mobiles.

La paroi de la bactérie étant en première ligne dans la perception et l'interaction de la bactérie avec son environnement, j'ai dans un premier temps caractérisé des réseaux de régulation impliqués dans la présentation de composants de la paroi ainsi que dans la perception des modifications du taux de D-alanylation du LTA. Cette analyse a mis en évidence que le processus de D-alanylation chez *S. agalactiae* est contrôlé de manière complexe par deux boucles de rétroaction faisant intervenir deux systèmes de régulation CiaRH et DltRS, permettant un ajustement fin de l'expression des gènes codant pour les protéines impliquées dans ce processus. À ce jour, cette double régulation semble unique à *S. agalactiae* parmi les streptocoques. En collaboration avec différentes équipes, nous avons aussi étudié les systèmes de régulation associés à l'expression de la protéine de virulence Srr et du pilus PI-2a et démontré une régulation croisée entre des régulateurs appartenant aux deux îlots génomiques SecA2/Srr et PI-2a, ainsi que le contrôle indirect de l'expression des gènes de ces îlots par un régulateur appartenant au core génome, CovRS. Enfin, au travers de ces différents projets j'ai pu réaliser une analyse combinée de 26 transcriptomes qui a conduit à la prédiction de cibles directes pour plusieurs régulateurs transcriptionnels chez *S. agalactiae*.

Dans un second temps, j'ai étudié la diversité d'une nouvelle famille d'éléments intégratifs et conjugatifs (ICEs) les TnGBS, impliqués dans des phénomènes de transfert de longs fragments d'ADN chromosomique par conjugaison, qui paraissent avoir joué un rôle important dans l'évolution du génome de *S. agalactiae*. Nous avons montré que cette famille possède deux particularités : il s'agit de la première famille d'ICE dont la mobilité dépend d'une transposase à motif DDE et non d'une intégrase; les TnGBS s'insèrent spécifiquement au niveau des régions promotrices. L'étude de la diversité des éléments de cette famille montre qu'ils se sont largement répandus au sein du genre streptocoque. Ils peuvent être groupés en deux sous-familles TnGBS1 et TnGBS2, qui associent une même transposase avec deux modules de conjugaison différents et qui présentent des caractéristiques de répllication et de régulation distinctes. L'analyse comparative des fonctions potentiellement portées par TnGBS1 et de TnGBS2 a permis de mieux définir les protéines impliquées dans les différentes phases du processus de conjugaison: transposition, mobilisation, transfert et répllication.