Inês Filipa Palão Silvestre

Mestre em Microbiologia Médica

Contribution of genomics and transcriptomics to the understanding of the biological role of DNases in *Streptococcus agalactiae*

Dissertação para obtenção do Grau de Doutor em
Biologia

Orientadora: Professora Doutora Maria José Gonçalves Gaspar Borrego, Investigadora Auxiliar, Instituto Nacional de Saúde Dr. Ricardo Jorge

Co-orientadora: Professora Doutora Rita Gonçalves Sobral de Almeida, Professora Auxiliar, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Júri:

Presidente: Doutora Isabel Maria Godinho de Sá Nogueira, Professora Catedrática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Arguentes: Doutor Arsénio do Carmo Sales Mendes Fialho, Professor Associado com Agregação, Instituto Superior Técnico, Universidade de Lisboa

Doutora Susana Margarida Paraíso Ferreira, Investigadora Auxiliar, Centro de Investigação em Ciências da Saúde da Universidade da Beira Interior

Vogais: Doutora Gabriela Conceição Duarte Jorge da Silva, Professora Associada, Faculdade de Farmácia, Universidade de Coimbra

Doutora Isabel Maria dos Santos Leitão Couto, Professora Associada, Instituto de Higiene e Medicina Tropical, Universidade NOVA de Lisboa

Doutora Maria José Gonçalves Gaspar Borrego, Investigadora Auxiliar, Instituto Nacional de Saúde Doutor Ricardo Jorge

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro, 2020

Inês Filipa Palão Silvestre

Mestre em Microbiologia Médica

Contribution of genomics and transcriptomics to the understanding of the biological role of DNases in *Streptococcus agalactiae*

Agradecimentos

O desafio, que culminou nesta tese de Doutoramento, certamente não teria sido possível sem o apoio incondicional de todos os que me acompanharam neste processo. Aqui expresso algumas palavras de agradecimento que seguramente serão insuficientes na representação de toda a minha gratidão e reconhecimento.

À minha orientadora, Doutora Maria José Borrego, pelas palavras de incentivo, acompanhamento constante, simpatia e disponibilidade demonstradas.

À minha coorientadora, Professora Doutora Rita Sobral, por aceitar integrar este projeto, pelo apoio, simpatia e disponibilidade.

À Professora Doutora Ilda Sanches, que se disponibilizou desde o primeiro dia para tudo o que fosse necessário na formalização do processo de candidatura a este Doutoramento. Infelizmente, o seu papel como orientadora foi breve mas permanecerá sempre na minha memória, pelo rigor, empenho, determinação e entrega à ciência. Foi sem dúvida uma inspiração, e por esse motivo lhe dedico esta tese.

A todos os membros da Comissão de Acompanhamento de Tese (CAT), por todas as discussões científicas e sugestões que enriqueceram este trabalho. Em especial, ao Doutor João Paulo Gomes pela sua dedicação e disponibilidade constantes. A sua motivação e incentivo, bem como a sua precisão científica marcaram este percurso. Agradeço, ainda, à Professora Rosário Mato Labajos pela colaboração científica, disponibilidade e simpatia.

À Professora Doutora Barbara Spellerberg, Institute of Medical Microbiology and Hygiene (IMMH), Ulm University, Alemanha, pelas estirpes, gentilmente cedidas, da sua coleção.

À Doutora Maria Luísa Jordão pelo apoio incondicional, disponibilidade constante e partilha de conhecimentos. A sua precisão experimental e a disponibilização do laboratório para os ensaios de biofilmes foram imprescindíveis para a sua realização.

A toda equipa do Núcleo de Bioinformática do Departamento de Doenças Infeciosas (DDI) do INSA, pelos momentos de boa disposição e pelo ambiente descontraído, sempre aliados a um elevado rigor e profissionalismo, em particular:

À Doutora Alexandra Nunes, pelo apoio, realização da análise bioinformática e transmissão de conhecimentos nessa área que me permitiram realizar a análise de alguns dos resultados obtidos de forma independente;

Ao Doutor Vítor Borges, pelo apoio, disponibilidade e realização da análise bioinformática, mas também pelo seu contributo geral para este trabalho, através das suas críticas sempre construtivas – desde a realização dos ensaios experimentais até à análise dos resultados finais.

Ao INSA, em particular ao Departamento de Doenças Infecciosas (DDI) na pessoa do seu coordenador, Dr. Jorge Machado, pelas condições disponibilizadas para a realização do presente trabalho.

A todos os membros do Laboratório Nacional de Referência de Infeções Sexualmente Transmissíveis do DDI pelo acolhimento, simpatia, disponibilidade e companheirismo, em particular, às colegas de laboratório Dr.ª Dora Cordeiro e Dr.ª Ana Isabel Salas.

À Doutora Mónica Oleastro, Doutora Leonor Silveira, Doutora Maria Manuela Caniça, Doutora Maria Eugenia Ferreira, Doutora Vera Manageiro e ao Dr. João Carlos Rodrigues pela cedência de tiras de E-test® e de discos de antibióticos que possibilitaram a realização dos antibiogramas.

À Dr.ª Irene Matos pela disponibilidade e preparação de meios de cultura que possibilitaram a realização de muitos dos ensaios.

À Doutora Rita Ferreira pela sua disponibilidade no acompanhamento dos primeiros ensaios experimentais, pela sua precisão e rigor científico na transmissão de conhecimentos.

Agradecimentos

A toda a equipa da Unidade de Tecnologia e Inovação do INSA, em especial ao seu responsável, Doutor Luís Vieira, à Dr.ª Catarina Carneiro Silva, e à Dr.ª Sílvia Vaz Duarte, pela disponibilidade e realização das metodologias de Sequenciação de Nova Geração, as quais foram tão relevantes neste trabalho.

À Professora Doutora Ana Tenreiro da Faculdade de Ciências da Universidade de Lisboa, pelo tempo que me dedicou e por me ter concedido acesso ao uso do fluorímetro utilizado nos ensaios quantitativos de produção de DNases.

À Hemovida®, na pessoa do Dr. Francisco Soares, pela cedência do Plasma Humano Inativado, que permitiu a conclusão dos ensaios experimentais da formação de biofilmes, mas também pelo apoio e palavras de reconhecimento e incentivo.

Ao Hospital de Santarém, em especial ao Serviço de Patologia Clínica, na pessoa da sua Coordenadora, Dr.ª Teresa Geada, e a todos os meus colegas, pela compreensão e apoio.

Às minhas colegas da Hemovida® do Hospital CUF Santarém, Sónia Domingues, Alexandra Franco, Andreia Venda, Susana Gomes e Carolina Serrano, pela amizade, disponibilidade, apoio e palavras de encorajamento.

À minha colega de doutoramento Cynthia Barroco, pela empatia, companheirismo e partilha de conhecimentos ao longo deste caminho, que percorremos em simultâneo.

A todos os que sempre me incentivaram a perseguir os meus objetivos. À minha família, em particular aos meus pais, pelas palavras de apoio e incentivo. Aos meus amigos por me apoiarem, muitas vezes à distância, e por permanecerem na minha vida ainda que a minha disponibilidade e dedicação tenha sido, muitas vezes, inferior. Agradeço todo o vosso carinho, compreensão e transmissão de confiança permanentes. A todos agradeço que tenham sempre acreditado que seria possível.

"We must find time to stop and thank the people who make a difference in our lives."

John F. Kennedy

Resumo

Streptococcus agalactiae é a principal causa de pneumonia neonatal, sépsis e meningite. O clone hipervirulento III/ST17, tem sido associado à meningite neonatal de início tardio, tendo a atividade de DNases extracelulares e a formação de biofilmes sido apontados como fatores de virulência nessas estirpes. O objetivo da presente dissertação de doutoramento foi contribuir para a descodificação da base genómica de produção de DNases em estirpes de *S. agalactiae*, isoladas em humanos, com diferentes tropismos celulares (invasivo *vs* colonização).

A produção diferencial de DNases de um conjunto de estirpes pertencentes às linhagens ST17 e ST19 foi investigada através da análise do *"core-genome"* e do *"pan-genome"*. As estirpes ST17, exceto uma, apresentaram atividade nucleásica contrariamente às estirpes ST19, em que apenas uma revelou um fenótipo DNase(+). A estirpe ST17 DNase(-) exibiu uma alteração aminoacídica exclusiva na proteína NucA e a presença de um homólogo do tranposão TnGBS2.3, para além de um fago intacto sem homologia em *S. agalactiae.* A estirpe ST19 DNase(+) revelou um aminoácido alterado na nuclease GBS0609. A análise transcriptómica permitiu avaliar os níveis de expressão génica, por categoria funcional, da estirpe de referência *S. agalactiae* NEM316, durante a fase exponencial de crescimento; a exposição ao DNA humano não induziu alterações no transcriptoma.

Foram determinadas as condições ótimas para a formação de biofilmes em estirpes de *S. agalactiae* de ST17 e ST19, contribuindo para a padronização dos procedimentos experimentais do estudo de produção de biofilmes nesta espécie bacteriana. A digestão enzimática de biofilmes maduros, para as três estirpes mais produtoras de biofilmes, evidenciou que o componente predominante da matriz extracelular são as proteínas.

Globalmente, a presente dissertação de doutoramento contribui para o conhecimento sobre a produção das DNases extracelulares e a capacidade de formação de biofilmes das linhagens ST17 e ST19 de *S. agalactiae*, providenciando novos dados genómicos e transcriptómicos.

Palavras-chave: *Streptococcus agalactiae*, DNases Extracelulares, Formação de Biofilmes, Sequenciação Total do Genoma, RNA-seq

Abstract

Streptococcus agalactiae is the leading cause of neonatal pneumonia, sepsis, and meningitis. Strains III/ST17 emerged as a hypervirulent clone mostly associated with meningitis during late-onset disease. Production of extracellular DNases and biofilm formation have been proposed to explain the leading role of this clone in neonatal meningitis. The aim of this thesis was to shed some light on the genetic background of DNase production in *S. agalactiae* strains isolated in humans presenting diverse cell tropism (invasive *vs* carriage).

Genomic approaches were used to decipher the molecular basis of the differential production of DNases, through the analysis of the "core-genome" and the "pan-genome" of ST17 and ST19 strains. All ST17 strains, except one, displayed DNase activity which was observed in only one ST19 strain. ST17 DNase(-) revealed an exclusive amino acid change in NucA, a TnGBS2.3 homolog and an intact phage without homology in *S. agalactiae.* ST19 DNase(+) revealed an exclusive amino acid change alteration in nuclease GBS0609. A transcriptomic approach allowed the analysis of the level of expression, by functional category, of genes of *S. agalactiae* reference strain NEM316 at the exponential growth phase; exposure to human DNA did not affect the transcriptome.

The optimal conditions for *S. agalactiae* biofilm assembly were determined for ST17 and ST19 strains, contributing to the standardization of experimental procedures, thus allowing the comparison of results between different laboratories. In addition, the enzymatic digestion of mature biofilms for the three strongest biofilm producers, evidenced that proteins were the predominant component of the extracellular polymeric matrix.

Overall, the findings presented in this Ph.D. thesis may contribute for the knowledge on the production of extracellular DNases, and provide new insights into biofilm formation, genomics and transcriptomics for ST17 and ST19 *S. agalactiae* lineages.

Keywords: *Streptococcus agalactiae*, Extracellular DNases, Biofilm formation, Whole Genome Sequencing, RNA-seq

Table of contents

Figure Index

Table Index

aa – amino acid

ACOG – American College of Obstetricians and Gynecologists

Alp – Alpha-like family surface proteins

ATCC – American Type Culture Collection

BBB – Blood brain barrier

β-H/C – β-hemolysin/cytolysin

BLAST - Basic local alignment search tool

bp – base pair

BrpA – Biofilm regulatory protein A

CC – Clonal complex

CDC – Centers for Disease Control and Prevention

cDNA – Complementary deoxyribonucleic acid

CFU – Colony-forming unit

$cMLS_B$ – Constitutive macrolide-lincosamide-streptogramin B resistance

CDS – Coding sequence

CO_2 – Carbon dioxide

COG – Cluster of orthologous groups

CPS – Type-specific capsular polysaccharide

CRISPR – Clustered regularly interspaced short palindromic repeats

CV – Crystal violet

CWA – Cell wall-anchored proteins

DNA – Deoxyribonucleic acid

dsDNA – Double stranded deoxyribonucleic acid

ECM – Extracellular matrix proteins

EDTA – Ethylenediaminetetraacetic acid

eDNA – Extracellular DNA

ENA – European nucleotide archive

EOD – Early-onset disease

EPS – Extracellular polymeric substances

erm – Erythromycin ribossome methylase

EUCAST – European Committee for Antimicrobial and Susceptibility Testing

FCT – Fundação para a Ciência e a Tecnologia

FCT NOVA – Nova School of Science and Technology

Fbs – Fibrinogen-binding proteins

FPKM – Fragments per kilobase of CDS per million mapped reads

GAPDH – Glyceraldehyde-3-phosphate dehydrogenase

GBS – Group B *Streptococcus*

h – hour

H^+ – Hydron

hiFBS – Heat inactivated fetal bovine serum

HvgA – Hypervirulent adhesin

IAP – Intrapartum antibiotic prophylaxis

ICEs – Integrative and conjugative elements

IGR – Intergenic region

IGV – Integrative genomics viewer

IHP – Inactivated human plasma

$iMLS_B$ – Inducible macrolide-lincosamide-streptogramin B resistance

IMMH – Institute of Medical Microbiology and Hygiene

INSA – Instituto Nacional de Saúde Doutor Ricardo Jorge

Lmb - Laminin-binding protein

LOD – Late-onset disease

M – Millions

Mbp/Mb – Millions of base pairs

MBP – Moderate biofilm producer

mef – Macrolide efflux

mg – Milligram

MGEs – Mobile genetic elements

MH-F – Mueller Hinton agar supplemented with 5% defibrinated horse blood and 20 mg/liter of β-NAD

MIC – Minimum inhibitory concentration

min – minutes

ml – Millilitre

MLSA – Multilocus sequence analysis

MLST – Multilocus sequence typing

mm – millimeter

mRNA – Messenger ribonucleic acid

NBP – Non biofilm producer

NETs – Neutrophil extracellular traps

nm – Nanometre

nr – non-redundant

nt – nucleotide

NT – Non-typeable strains

NucA – Nuclease A

OD – Optical density

ODc – Cut-off value

ORF – Open reading frame

PBS – Phosphate-buffered saline

PCR – Polymerase chain reaction

pH – Hydrogen ion concentration

PI – Pilus islands

RBS – Ribosome binding site

RNA – Ribonucleic acid

RNA-seq – RNA-sequencing

rpm – Rotations per minute

RPS-BLAST – Reverse Position-Specific BLAST

rRNA – Ribosomal ribonucleic acid

RT – Room temperature

SAM – S-adenosylmethionine

SBP – Strong biofilm producer

SD – Standard deviation

SEM – Scanning electron microscopy

SNP – Single nucleotide polymorphism

SNV – Single nucleotide variant

Srr – Serine-rich repeat

ST – Sequence type

T4SS – Type IV secretion system

TCS – Two-component signal transduction system

TE – Tris-ethylenediaminetetraacetic acid

THB – Todd Hewitt Broth

tRNA – Transfer ribonucleic acid

U – Kunitz Unit

USA – United States of America

VFDB – Virulence factors database

WBP – Weak biofilm producer

WGS – Whole genome sequencing

wgMLST – Whole genome multi locus sequence typing

WHO – World Health Organization

The body of this Ph.D. dissertation is based on six chapters (listed below), including a general introduction (Chapter I) and a final overview (Chapter VI). Four chapters (II-V) are presented as individual manuscripts. As Chapter II is a manuscript written in a review article format, the general introduction (Chapter I) contains only a succinct overview of the *S. agalactiae* biology. One of the manuscripts (Chapter V) has been published in a peer reviewed international journal; Chapter III and IV have been published as preprints; Chapters II and III were submitted for publication in peer reviewed international journals at the time this thesis was completed. The chapters are organized following a rational order considering the objectives delineated for this Ph.D. work. Furthermore, they do not perfectly reflect the chronological order of the events, as several studies were developed simultaneously. Briefly, each chapter includes the following contents:

Chapter I – General Introduction

This chapter consists in a general introduction that provides an overview of *S. agalactiae*, starting with a global description of the major aspects of the biology and taxonomy, and a brief summary of the historical background. The clinical relevance, preventive strategies, antibiotic resistance, molecular epidemiology and pathogenesis are also reviewed in this chapter. Furthermore, it includes a brief overview of *S. agalactiae* genomics and transcriptomics. Finally, the main objectives of this Ph.D. thesis are detailed and contextualized.

Chapter II – Exploring the factors accounting for the epidemiologic success of *Streptococcus agalactiae* clone III/ST17

This chapter consists in a review of the genomic characteristics of *S. agalactiae* underlying the main virulence factors that contribute to the epidemiological success of the III/ST17 clone. The specific contribution of extracellular DNases and biofilm formation for capsular type III/ST17 hypervirulence are also considered.

Silvestre, I., Jordão L., Borrego M. J. Exploring the factors accounting for the epidemiologic success of *Streptococcus agalactiae* clone III/ST17 (*submitted manuscript*).

Chapter III – Genomic insights on DNase production in *Streptococcus agalactiae* ST17 and ST19 strains.

This chapter describes a comparative genomic analysis by Whole Genome Sequencing (WGS) of ST17 and ST19 *S. agalactiae* human strains with different cell tropism (invasive vs carriage) and diverse DNase production phenotypes. The study provides the identification of genomic features that potentially contribute to DNase activity although these data require further research to test their robustness, namely by evaluating *S. agalactiae* strains of different ST and

diverse DNase production phenotypes, in order to reach a better understanding of the pathways through which DNases contribute to *S. agalactiae* pathogenesis.

Silvestre, I., A. Nunes, V. Borges, J. Isidro, C. Silva, L. Vieira, J. P. Gomes, and M. J. Borrego. 2020. Genomic insights on DNase production in *Streptococcus agalactiae* ST17 and ST19 strains. BioRxiv. DOI: 10.1101/2020.12.09.418327

Silvestre, I., A. Nunes, V. Borges, J. Isidro, C. Silva, L. Vieira, M. J. Borrego, and J. P. Gomes. Genomic insights on DNase production in *Streptococcus agalactiae* ST17 and ST19 strains (*submitted manuscript*).

Chapter IV – Global gene expression analysis of *Streptococcus agalactiae* at exponential growth phase

This chapter consists in a first *in vitro* attempt to systematize the expression levels of *S. agalactiae* reference strain NEM316 at the exponential growth phase by RNA-sequencing (RNA-seq). Preliminary assays to understand whether direct human DNA exposure affects *S. agalactiae* gene expression evidenced no effect.

This chapter corresponds to the following manuscript deposited in BioRxiv: Silvestre, I., V. Borges, S. Duarte, A. Nunes, R. Sobral, L. Vieira, J. P. Gomes, and M. J. Borrego. 2020. Global gene expression analysis of *Streptococcus agalactiae* at exponential growth phase. BioRxiv. DOI: 10.1101/2020.11.13.381939.

Chapter V – Biofilm formation by ST17 and ST19 strains of *Streptococcus agalactiae*

This chapter corresponds to the following published manuscript: Silvestre, I., M. J. Borrego, and L. Jordão. 2020. Biofilm formation by ST17 and ST19 strains of *Streptococcus agalactiae*. Research in Microbiology. 171(8):311-318. DOI: 10.1016/j.resmic.2020.08.001. The aim of this *in vitro* study was to evaluate the effects of three environmental factors (H^+, glucose and human plasma) in biofilm formation, by carrier and invasive *S. agalactiae* strains of ST17 and ST19, DNase producers and non-producers. The composition of the extracellular polymeric matrix was accessed for ST17 strains with the strongest biofilm production, and proteins were shown to be the predominant component.

Chapter VI – Final overview, concluding remarks and future perspectives

This chapter provides a global overview of the subjects addressed throughout the previous chapters, reviewing the main results and the conclusions achieved through this Ph.D. thesis. Unsolved questions that should be addressed in a future follow-up are also presented.

Considering the dissimilar layouts and in-text reference styles adopted by the different journals to which the manuscripts were submitted or published, all chapters are formatted in a unique style, and a single section of "References" is presented with references sequentially numbered according to the order in which they appear in the text. Of note, all abbreviations were also uniformized. Finally, supplementary data of each chapter is also compiled in a single section designated by “Supplementary Material”. Whenever it is impossible to present data in a printable format, a link to the online access is provided.

Chapter I

General Introduction

1. General Introduction

1.1. *Streptococcus spp.*

The first record of *Streptococcus*, although not named as such, occurred in 1683 in van Leeuwenhoek's drawings of microscope images. However, the main entry of streptococci into history happened in 1879, when Louis Pasteur was studying puerperal fever, which was causing high mortality rates in maternity wards. Louis Pasteur found rounded granules (microorganisms) arranged in the form of chains or strings of beads in the body of diseased women. It was later recognized that these granules were the cause of infections contracted by those women after giving birth [1].

The term *Streptococcus* comes from the Greek strepto (twisted) and coccus (spherical). The earliest attempt to differentiate streptococci was probably made in 1903 by Shottmuller, who used blood agar to differentiate strains according to the type of hemolysis [2]. Streptococci could be then classified into β-hemolytic, α-hemolytic or γ-hemolytic depending on the erythrocyte's hemolysis, that could be complete, partial, or not induced, respectively [3].

Streptococci are spherical, ovoid or lancet-shaped (0.5 and 2 μm in diameter), Gram-positive cocci (Figure 1.1), catalase negative, non-motile and non-spore forming. Since cellular division of *Streptococcus spp.* occurs along a single axis or plane, these bacteria grow in pairs or chains (Figure 1.1). After 18–24 h of incubation at 35–37 °C on blood agar, typically grayish-white, smooth, glossy, and translucent colonies appear, with zones of α/β-hemolysis or no hemolysis. Some species require CO_2 for initial isolation, but strains may lose this requirement in subculture [3-5].

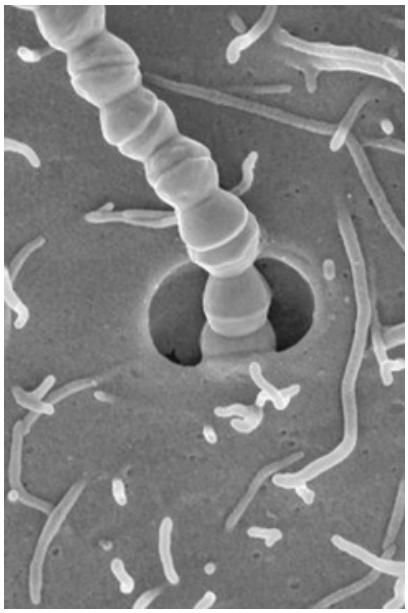


Figure 1.1 Electron micrograph of streptococci invading a human cell (reproduced from Quah *et al.,* 2017).

In 1934, Rebecca Lancefield implemented the so-called 'Lancefield scheme' for streptococci classification [6], which groups streptococcal strains according to the carbohydrate composition of the antigens present in the cell wall [1-4, 7]. So far, thirteen different serological groups have been identified, out of which groups A, B, C, and G, *Streptococcus pneumoniae,* and streptococci of the viridans group are the most important regarding human health [7]. Thus, although various parameters and methods can be used to differentiate streptococci (e.g. colony size, hemolysis, fermentation ability and tolerance tests), agglutination tests to determine the Lancefield group remain an important routine technique for the identification of beta-hemolytic streptococci, because of their simplicity, rapidity, and specificity [8, 9]. Its widespread application is hindered by the fact that group-specific antigens for other species may be absent or shared between distinct taxa, and this constitutes a limitation of the methodology [1, 2, 7].

The availability of modern molecular techniques brought many changes to the taxonomy of the *Streptococcus* genus during the last decade; in fact, molecular methods such as 16S rRNA gene sequencing, multilocus sequence analysis (MLSA) and WGS analysis allowed to determine the taxonomic relationships between streptococcal species [7]. Based on 16S rRNA gene sequences, six groups of *Streptococcus* were defined: Anginosus, Pyogenic, Bovis, Salivarius, Mutans and Mitis, among which the pyogenic group included *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus equi*, *Streptococcus dysgalactiae*, *Streptococcus uberis*, and *Streptococcus iniae* [1]. Later, comparative genomic analyses of the genus *Streptococcus* indicated that there are two major evolutionary lineages, with Pyogenic, Bovis, Salivarius and

Mutans groups forming one population, while Mitis, Anginosus and unknown groups clustered into another (Figure 1.2) [10].

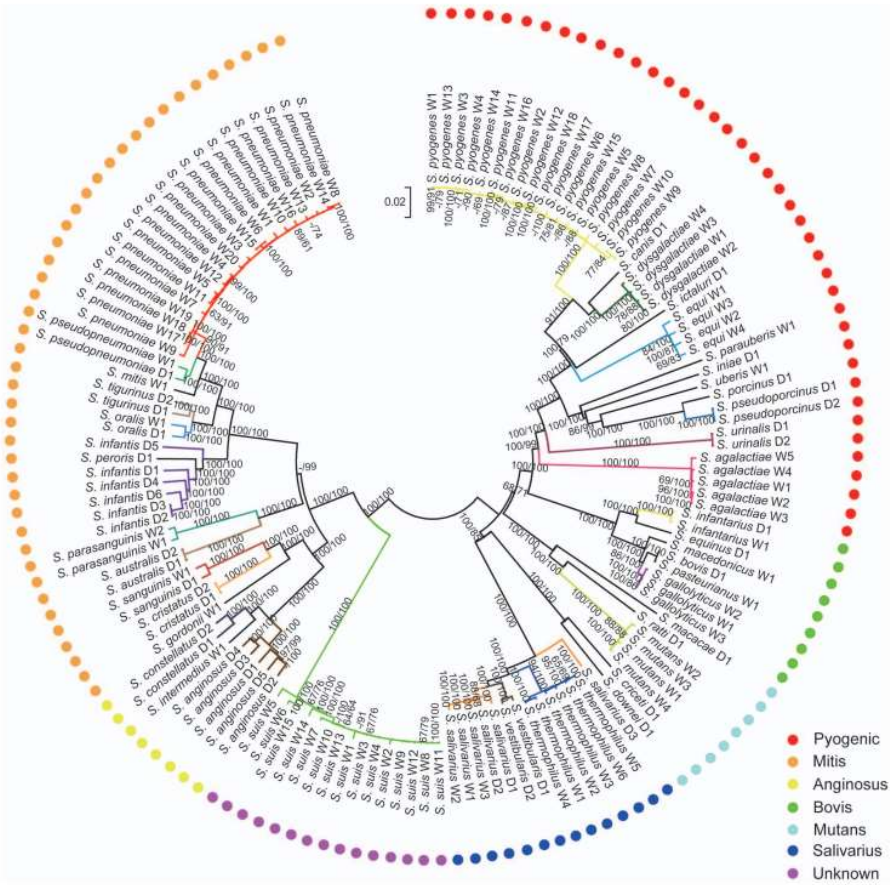


Figure 1.2 Phylogenomic tree of *Streptococcus*. Different color-coded branches denoted different species (reproduced from Gao *et al.*, 2014).

1.2. *Streptococcus agalactiae*

S. agalactiae is equally referred to as group B *Streptococcus* (GBS) as it is the only streptococcal species harboring cell-wall-specific polysaccharide antigen of the Lancefield group B, which consists of repeated units of rhamnose, and is common to all *S. agalactiae* strains [11]. In most laboratories, diagnosis is based on colonies displaying typical *S. agalactiae* morphology that are subjected to latex agglutination tests to determine the presence of the Lancefield group B antigen. However, *Streptococcus porcinus*, which may be present in the genital tract of pregnant women, also grows on blood agar as β-hemolytic colonies and agglutinates with *S. agalactiae* latex agglutination tests. For this reason, the detection of beta-hemolytic colonies that are positive in *S. agalactiae* latex agglutination tests require further confirmation [9].

In recent years, several specific Polymerase Chain Reaction (PCR) assays have been developed for the detection of *S. agalactiae* colonization, employing different genes as targets;

the CAMP factor gene was the first to be used [9, 12]. The main advantage of PCR is the short time necessary to provide results regarding *S. agalactiae* colonization status at the point of delivery. However, even PCR testing could take too long, preventing the implementation of intrapartum antibiotic prophylaxis (IAP), as many women deliver within a few hours after hospital admission [9]. Currently, most of the nucleic acid amplification tests (NAATs) are incorporated into automated molecular platforms such as the BD MAX^{TM} GBS system (Becton, Dickinson), the Smart Cycler® (Cepheid), and specially the Xpert® technology (Cepheid), which provides results in less than one hour. Nonetheless, in comparison with culture enrichment, PCR is considered to have lower sensitivity and to lacks to provide information on antibiotic susceptibility, which is required for all women reporting allergy to penicillin [9].

A correct and reliable species identification can be easily achieved through the 16S rRNA gene and *sodA* gene sequencing [9]; however, WGS is the gold standard for taxonomic purposes.

1.2.1. Historical background

S. agalactiae was recognized in 1920 as the etiologic agent of bovine mastitis and firstly identified in humans in 1935, by Lancefield and Hare, on vaginal secretions of asymptomatic women; accordingly, it was considered as a colonizing agent [13]. In 1938, Fry reported for the first time the presence of *S. agalactiae* in three fatal cases of puerperal sepsis in women [14].

In the early 1960s, the association between colonization of the maternal genital tract by *S. agalactiae* and disease in the newborn was established [4, 15, 16]. Until 1965, occasional cases of severe *S. agalactiae* disease were reported but, during the 1970s, *S. agalactiae* came to be considered as the leading cause of septicemia and neonatal meningitis in geographically diverse regions [17-21]. *S. agalactiae* emerged as an important pathogen for humans, responsible for invasive infections during the first week of life of the newborn, with mortality rates of 20-50% [4, 11, 22-24]. An increase in the frequency of isolation of this microorganism from neonates was uncovered and *S. agalactiae* came to be considered as the most common cause of meningitis due to Gram-positive bacteria during the first three months of life, ranked second only to *Escherichia coli* as a cause of neonatal meningitis [18, 25].

1.2.2. Clinical Relevance

S. agalactiae, a common inhabitant of the healthy human gastrointestinal and urogenital tract, is an opportunistic pathogen in animals and humans, being the leading cause of human neonatal pneumonia, sepsis and meningitis [26]. Annually, 2.7 million infants are estimated to

die due to perinatal health complications, of which at least 90,000 are due to invasive *S. agalactiae* disease [27, 28].

The onset of *S. agalactiae* infections takes place usually within the first week of life, being designated as early-onset disease (EOD). During uterine life, the contamination of the amniotic fluid by *S. agalactiae* may occur due to the rupture or invasion of the membranes and subsequent colonization of the fetal respiratory tract. However, most infections occur during delivery through aspiration of infected amniotic fluid or direct contact with the colonized vaginal mucosa of the mother, causing pneumonia, sepsis and, less commonly, meningitis [4, 22, 29-31]. *S. agalactiae* infections may happen a little later, in infants aged over one week, with most infections becoming evident during the first three months of life, and causing late-onset disease (LOD) [22, 32]. In LOD, transmission of *S. agalactiae* occurs through horizontal transmission by maternal, nosocomial, or community sources, leading to bacteremia and meningitis [4, 22, 24, 33, 34]; regardless of whether it is a consequence of EOD or LOD, meningitis in infants will surely induce significant cognitive or neurological sequelae [30, 35]. Thus, asymptomatic rectovaginal colonization among pregnant women is considered a public health problem, because it is recognized as the primary risk factor for disease development [22, 35]. *S. agalactiae* maternal colonization may be intermittent throughout pregnancy; overall, it was estimated at 18% [36], but varies worldwide, with rates from 10 to 30% in the United States of America (USA), 6.5% up to 36% in Europe, 7.1% to 16% in Asia, 9.1 to 25.3% in Middle East, and 11.9 to 31.6% in Africa [37]. Other maternal risk factors for EOD include preterm delivery, prolonged (>18 h) rupture of membranes, intrapartum fever (>38 °C), young maternal age, EOD by *S. agalactiae* in previous delivery or maternal black race [15, 22-24, 38].

During pregnancy, *S. agalactiae* may be responsible for urinary tract infection, chorioamnionitis, endometritis, sepsis, meningitis, septic pelvic thrombophlebitis, premature rupture of membranes, preterm births, low-birth-weight newborns or miscarriage [9, 15, 24, 30, 39]. The risk of transmission in the case of maternal colonization was established at 30-70% and, of those colonized neonates, about 1-2% is expected to develop EOD [37]. A fatality rate of 2 to 10% was estimated for EOD, and fatal outcomes seem to be more frequent among premature neonates [9]. The mortality rate for LOD is lower, but meningitis and subsequent sequelae are more frequently associated with LOD, rather than with EOD [9].

More recently, experts recognized the increasing impact of invasive *S. agalactiae* disease in adults, particularly in persons with chronic disabilities, the elderly and otherwise immunocompromised individuals, among who it causes bacteremia, urinary tract infection, endocarditis, skin and soft tissue infections, arthritis, pneumonia, osteomyelitis and, less frequently, meningitis [4, 22, 24, 30, 35, 40].

1.2.3. Prevention of neonatal infection

The first reports showing that administering IAP to mothers who are colonized with *S. agalactiae* (and who are typically asymptomatic) reduces vertical transmission, and prevents *S. agalactiae* EOD, were published in the 1980s [9]. These findings led to the publication of the first guidelines in 1996 by the Centers for Disease Control and Prevention (CDC), later updated and republished in 2002 and in 2010 [22]. The screening for *S. agalactiae* by collecting lower vaginal and rectal swabs at 35-37 weeks of pregnancy and intrapartum treatment of pregnant carriers, came to be considered fundamental for the prevention of neonatal infection [9, 15, 24, 29, 41]. The successful administration of IAP and the treatment of severe *S. agalactiae* infections rely on efficient and reliable detection of *S. agalactiae* in clinical samples [9]. Because the correlation between isolation in the last trimester of pregnancy and the time of delivery is not clear (since colonization with *S. agalactiae* may be transient, intermittent or persistent), the timing of specimen collection is important [9, 15, 24, 29, 41]. A weak concordance between prenatal and intrapartum culture results could be attributed to the time of prenatal *S. agalactiae* screening, but also to laboratory methodologies and antibiotic usage [42].

In Portugal, a national epidemiological survey performed between 2001 and 2005 revealed an incidence of EOD of 0.44 per 1,000 live births and similar mortality rates for EOD and LOD [43]. Worldwide, the widespread implementation of CDC guidelines has resulted in significant reductions in the incidence of EOD (USA: from 1.7 to about 0.37 per 1,000 live births between 1993 and 2008, Spain: from 2.4 to 0.45 per 1,000 live births between 1996 and 2008) [9, 22-24].

In 2020, the American College of Obstetricians and Gynecologists (ACOG) provided a Committee Opinion with an update to the recommended prophylaxis and prevention strategies for women during pregnancy and labor. It proposes *S. agalactiae* universal screening at 36 weeks of gestation, providing a 5-week window for valid culture results up to the gestational age of 41 weeks (likely reducing the reported incidence of discrepant antepartum culture results *vs* colonization at the time of birth) [38].

Nonetheless, IAP has no impact on LOD, which continues the leading cause of neonatal morbidity and mortality [24, 37, 44]. In addition, newborn exposure to antibiotics has been associated with alterations in the gut microbiome, leading to subsequent allergies, asthma and obesity [38]. Moreover, IAP can hardly be implemented in many low- and middle-income countries [26]. Thus, the development of vaccines was considered a priority by the World Health Organization (WHO), for maternal immunization against *S. agalactiae*, particularly in low- and middle-income countries [28].

1.2.4. Intravenous Antibiotic Prophylaxis Intrapartum

S. agalactiae is sensitive to β-lactam antibiotics, especially to penicillin and ampicillin [15, 45]. Intravenous penicillin remains the treatment of choice for infections caused by *S. agalactiae* in adults [15, 38, 46], with ampicillin being considered an acceptable alternative [38]. The use of penicillin intrapartum relates to its effectiveness in transplacental passage and its low cost [15].

For women reporting penicillin allergy and low risk of anaphylaxis, first-generation cephalosporins (i.e., cefazolin) are recommended [38]. Macrolides (erythromycin) and lincosamides (clindamycin) have also emerged as alternative therapies for patients allergic to penicillin [15, 38, 47]. However, rates of *S. agalactiae* resistance to erythromycin have been increasing (up to 44.8%) and, in addition, erythromycin may not reliably reach the amniotic fluid and fetal tissues; as such, the use of erythromycin is no longer recommended [22, 38]. Currently, for women allergic to penicillin with high risk of anaphylaxis, clindamycin administration is recommended, but antimicrobial testing is required to determine whether the *S. agalactiae* strain is susceptible to clindamycin or not [22, 38]. Thus, intravenous vancomycin remains the only pharmacokinetically and microbiologically validated option for IAP in women: a) who report a high-risk penicillin allergy; b) whose *S. agalactiae* strain is resistant to clindamycin; c) with unknown susceptibility to clindamycin and erythromycin [22, 38].

1.2.5. Antibiotic resistance

The widespread use of IAP to prevent *S. agalactiae* EOD has raised concern about the development of antibiotic resistance, justifying the need to constantly monitor the *S. agalactiae* susceptibility patterns to antimicrobials [22, 30, 45, 48]. The metabolic features underlying resistance to antibiotics in *S. agalactiae* include antibiotic inactivation (by enzymes), efflux pumps (cell antibiotic expelling before it reaches its target and produces an effect), and modification of the antimicrobial agent binding site (i.e. by acquisition of genetic material, deletion, or mutations) [49, 50].

While the proportion of isolates resistant to erythromycin and clindamycin has increased in recent years [51], cases of resistance to penicillin or ampicillin have not been described; nonetheless, some *S. agalactiae* isolates exhibit reduced susceptibility to penicillin and ampicillin [22, 46, 52-54]. In Portugal, an increasing rate of *S. agalactiae* resistance to macrolides was detected in a study carried out between 2009 and 2015 (35.1% and 33.9%, respectively) [55]. The

resistance to macrolides occurs primarily through an efflux mechanism, mediated by a macrolide efflux protein encoded by *mefA*/*mefE* genes [56, 57], the so-called M phenotype [47]. Another resistance mechanism is a ribosomal modification mediated by an erythromycin ribosome methylase encoded by *ermA* (subclass *ermTR*) and *ermB* genes [56,57]. This modification consists in the methylation of the 23S rRNA region that alters the binding-target for the antibiotic, leading to a MLS_B phenotype and to a cross-resistance to macrolides, lincosamides and streptogramin B [47, 58]. This resistance can be expressed constitutively ($cMLS_B$, constitutive macrolide-lincosamide-streptogramin B resistance) or induced by erythromycin ($iMLS_B$, inducible macrolide-lincosamide-streptogramin B resistance) [47, 56, 57, 59, 60]. Other genes, like *msr(D)*, could contribute for the expression of erythromycin resistance. Studies carried out in pneumococci demonstrated that *mef* and *msr(D)* are cotranscribed, suggesting that the proteins encoded by the two genes may act as a dual efflux system, inducible by erythromycin. However, it has also been suggested that the efflux system encoded by *msr(D)* is capable to function independently of the one encoded by *mef* [61].

The first vancomycin-resistant *S. agalactiae* strain was described in 2013. Vancomycin and other glycopeptides bind to D-alanine-D-alanine termini of peptidoglycan precursors, which prevents transglycosylation and transpeptidation reactions of peptidoglycan synthesis. Vancomycin resistance is conferred by the *vanG* operon through the synthesis of peptidoglycan precursors with C-terminal D-alanine-D-serine residues that have low affinity towards vancomycin. The collective observations made from *S. agalactiae* strains resistant to vancomycin revealed a specific hot spot for insertional elements that is conserved between streptococci and other Gram-positive species; these strains potentially represent a *S. agalactiae* lineage that is predisposed to insertion of *vanG* elements [62, 63].

Fluoroquinolones emerged as an important alternative for the treatment of adult *S. agalactiae* infections [64]. However, *S. agalactiae* strains that are highly resistant to fluoroquinolones have been reported in different geographical locations [65, 66, 67]. Fluoroquinolones inhibit bacterial growth primarily by binding to enzymes involved in DNA replication, DNA gyrase and DNA topoisomerase IV. Specific mutations found in a region (quinolone resistance-determining region) of the *par* and *gyr* genes alter the amino acid composition of these enzymes and result in decreased binding and activity [66].

Finally, although tetracycline is not used to treat *S. agalactiae* infections, high rates (> 80%) of tetracycline resistant *S. agalactiae* strains have been described [55, 68-70], mostly associated with the *tet(M)* gene, despite *tet(O)* and *tet(L)* have also been identified [55, 69, 70]. The resistance to tetracycline can be the result of an efflux mechanism (mediated by *tet(K)* and *tet(L)* genes), protection of the ribosomal binding site (mediated by *tetM* and *tetO* genes) or changes in the permeability of the cell envelope that reduce the drug uptake [71].

1.2.6. Molecular Epidemiology

The *S. agalactiae* capsular polysaccharide (CPS, type-specific capsular polysaccharide) has been used to differentiate *S. agalactiae* strains by serotyping, particularly in epidemiological studies [72, 73]. The structure of the capsular polysaccharide of *S. agalactiae* is determined by the *cps locus*, which encodes for enzymes responsible for the synthesis of polysaccharides, recognized by precipitation methods or latex agglutination [73-76]. Thus, serotyping allowed classification of *S. agalactiae* strains in capsular polysaccharide-based serotypes Ia, Ib and II to IX [9, 32, 77, 78]. However, there are some non-typeable strains (NT), which probably do not express the capsular polysaccharide due to insertions or mutations in the capsular genes [76, 79-81].

Amplification and sequencing of the *S. agalactiae cps* cluster allowed to determine the correlation between capsular serotypes and genotypes and, although some strains remain NT by molecular genotyping, this methodology reduced the number of NT strains [76, 79]. Molecular genotyping is highly reproducible and discriminatory, resolving capsular-serotyping discrepancies and identifying new genotypes [79, 82]. The *cps* cluster comprises *cpsA-O*, *cpsR*, *cpsS*, *cpsY* genes, which exist and are conserved in most of the strains. However, *cpsG-K* genes are highly variable in genotypes Ia, Ib and II to VII, while *cpsE-cpsK* genes are variable in the genotype VIII [83, 84].

Capsular-type distribution varies by geographic region and ethnic origin [37, 57, 79, 85]; in fact, Ia, Ib, II, III, and V are often associated with colonization in the USA and Europe [37] while VI and VIII were identified as the most prevalent among pregnant women in Japan [15, 37, 79, 81, 86]. In Portugal, capsular-types Ia, III and V were identified as the most frequent in asymptomatic colonization [32, 87, 88] while the Ia and III were more frequent in cases of invasive disease [88]. Capsular-types VIII and IX were recently identified in Portugal among neonatal invasive disease cases, suggesting their recent introduction in the country [32, 55]; however, *S. agalactiae* capsular-type IX was first reported back in 2007, in Denmark [76].

S. agalactiae capsular-type III is considered the most dominant invasive clone worldwide, being responsible for most cases of neonatal meningitis [36, 37, 44, 85, 87, 89, 90], while capsular-types Ia [37] and V are more common in non-pregnant adults [36-37, 57]. Capsular-type IV has become more frequent and has been considered as an emerging pathogen for neonates and adults similarly to type V in the 1990s; curiously, association of capsular-type IV with HIV infections has been reported [87, 91-97]. In Portugal, a remarkable increase in capsular-type IV frequency has been observed, from 1% in 2006 to 20% in 2012, ranking it as the second most detected capsular-type in 2012 [87].

Multilocus sequence typing (MLST) contributes for the recognition of genetic lineages within capsular types that were shown to differ in virulence and tropism [98, 99]. The MLST methodology has been used in the molecular characterization of *S. agalactiae* strains allowing to define the population structure by identifying genetic lineages [99, 100]. In *S. agalactiae*, MLST is a typing method involving the sequencing of fragments with approximately 500 base pairs (bps), of seven housekeeping genes (*adhP, pheS, atr, glnA, sdhA, glcK, tkt*) which belong to the stable "core" of the bacterial genome ("core-genome") [96, 99, 100]. The combination of these seven *loci* alleles assigns each strain with a single allelic profile or sequence type (ST). The criterion for grouping different STs in a clonal complex (CC) is based on the number of alleles shared by several STs [99, 101]. One ST can include strains from different capsular-types. For this reason, it is assumed that the genes that encode the enzymes for the capsular biosynthesis in *S. agalactiae* are subjected to relatively frequent horizontal transfer, as described for *S. pneumoniae* [74, 96, 99, 100].

With the availability of MLST, it became clear that *S. agalactiae* strains of certain CCs possess a higher potential to cause invasive disease, while other CCs harbor mainly colonizing strains [37]. MLST of *S. agalactiae* isolated from different countries showed that most human carriage and invasive strains cluster into only five major CCs (CC1, CC10, CC17, CC19 and CC23); a sixth CC, CC26, is common in African countries [69]. The dominance of strains belonging to CCs 1, 19 and 23 among asymptomatic pregnant woman has been consistently reported; CC1 is also related with adult invasive infection [37]. *S. agalactiae* CC17 is the prototype of an emerging clone that successfully spread worldwide; in fact, the increased prevalence of neonatal invasive infection observed from the 1960s was strongly associated with this hypervirulent clone, almost exclusively composed by capsular-type III strains [90, 102, 103].

Genotypic tools such as CRISPR (clustered regularly interspaced short palindromic repeats) have been described in *S. agalactiae* and allow a fine characterization of the *S. agalactiae* population structure [104, 105].

1.2.7. Pathogenesis

S. agalactiae colonization and persistence in different host niches relies on the adherence capacity (via adhesins) to host cells, stress response mechanisms, defenses against other microorganisms, and immune evasion strategies [34, 37, 106, 107]. The factors mediating *S. agalactiae* transition from commensal to invasive pathogen – resulting in sepsis or meningitis – are not well understood, but they should derive from both the bacteria and the host [106].

S. agalactiae ability to adhere to host cells and extracellular matrix proteins (ECM), such as collagen, fibrinogen, fibronectin and laminin, were revealed to be crucial for both colonization and invasion of host barriers [34, 37, 106]. To mediate adherence to ECM and host cell surface, *S. agalactiae* expresses several surface-associated protein families including sets of cell wall-anchored (CWA) proteins, lipoproteins, and a variety of secreted proteins [108]. Adhesins facilitate persistence and intimate contacts between the bacterium and the host. *S. agalactiae* has many mechanisms to adapt and survive in its host and these mechanisms are often controlled via two-component signal transduction system (TCS), mediating the transition to invasive infections [37, 109]. In fact, 17-21 TCS critical to control virulence, adherence, resistance to host defenses, and bacterial metabolism were identified in *S. agalactiae* [110]. This suggests that *S. agalactiae* strains have a higher capacity to respond to a multitude of environment conditions, than other important human streptococcal pathogens, like *S. pneumoniae* (~14 TCS) and *S. pyogenes* (~13 TCS). The ability of *S. agalactiae* to survive whether in the acidic vaginal environment or within the blood should be largely mediated by the two-component regulator CovR/S, which is considered to regulate ~27% of the entire genome [37, 107, 110]. CovR/S seems to impact on virulence gene expression, including downregulation of fibrinogen-binding proteins (Fbs), multiple non-pilus adherence factors, genes involved in iron uptake, and the *cyl* operon implicated in production of the β-hemolysin/cytolysin (β-H/C) toxin and the antioxidant carotenoid pigment [34].

1.2.8. Genomics and transcriptomics

WGS analysis was applied to study *S. agalactiae* evolution as well as phylogenetic relationships between isolates from various hosts and sources [70]. The first *S. agalactiae* sequenced genomes were from NEM316 (III/ST23) and 2603V/R (V/ST110) [11, 111], followed by six additional human strains, namely COH1 (III/ST17) [112]. Actually, more than one hundred human *S. agalactiae* complete genomes are available (https://www.ncbi.nlm.nih.gov/genome/). The analysis of *S. agalactiae* genomes have shown a similar size (~ 2 Mbp), number of genes per strain (~ 2000 predicted coding regions), and GC content (~ 35%).

The identification of several new genes per sequenced genome reflects the genomic diversity of *S. agalactiae* (*open pan-genome*) [11, 111, 112]. *S. agalactiae* strains may be considered as displaying a *pan-genome* that consists in a *core-genome* shared by all strains (representing 80% of the genome) and a specific *acessory-genome* of each strain [11, 111, 112]. Many of these acquired specific genes have been associated with mobile elements, including bacteriophages, transposons and insertion sequences [111]. The acquired genes tend to locate in

chromosomic regions of the *accessory-genome* called genomic islands of pathogenicity, and contribute to *S. agalactiae* genetic diversity and adaptation to specific ecological niches, impacting on the emergence of this bacterium as an important pathogenic agent [111, 112]. Comparative genomic analysis has shown great similarity between the genomes of *S. agalactiae* and other streptococci, namely *S. pyogenes* and *S. pneumoniae*, supporting the possibility of acquisition of virulence traits from other species [111-113].

The transcriptome analysis is a comprehensive approach to the analysis of all the RNA transcripts in a specific physiological condition, playing an important role in the detection of gene expression and transcriptional regulation [114]. The understanding of bacterial transcription has greatly advanced due to microarrays, which quantify a set of predetermined sequences, and RNA-seq, which uses high throughput sequencing to captures all sequences [115]. RNA-seq allows for differential-expression analysis and it can also help to infer the functions of previously unannotated genes. RNA-seq technology has been successfully applied to clarify some molecular mechanisms underlying pathogenesis in *S. agalactiae* [110, 116-121], like the adaptation to the bovine environment or the survival in human whole blood [116, 120]; however, the transcriptional networks that coordinate the expression of genes involved in host colonization, virulence, or response to different host environments, remain largely unknown [117, 118].

1.3. Objectives

The major goal of this thesis was to shed some light on the genetic background that supports the production of extracellular DNases by *S. agalactiae* strains from different clinical scenarios, and to study their functional role during the initial steps of pathogenesis, namely on the production of biofilms. In order to achieve this objective, we carried out several studies.

In detail, the following objectives were pursued, and constitute the subject of each chapter:

i) Phenotypic characterization regarding extracellular DNase production and genomic characterization through WGS of ST17 and ST19 *S. agalactiae* strains (n=33) from human origin, with different cell tropism (invasive vs carriage). The clinical strains were selected among an enlarged set involving the collections of the Portuguese National Institute of Health (Instituto Nacional de Saúde Doutor Ricardo Jorge, INSA, IP) and the Institute of Medical Microbiology and Hygiene (IMMH), Ulm University, Germany;

ii) *In vitro* determination of the expression levels of *S. agalactiae* reference strain NEM316 during the exponential phase of growth with no inflicted environmental stress. In addition, disclosure of genes that are up and downregulated upon incubation with (*vs* without) human DNA (used to mimic a NET environment);

iii) Evaluation of the *in vitro* effects of three environmental factors (H^+, glucose and human plasma) in biofilm formation, by carrier and invasive *S. agalactiae* strains belonging to ST17 and ST19, including DNase producers and non-producers. The composition of the extracellular polymeric matrix was also accessed for ST17 strains with the strongest biofilm production phenotype.

Chapter II

Exploring the factors accounting for the epidemiologic success of *Streptococcus agalactiae* clone III/ST17

The contents of this chapter were submitted (with minor changes) to a peer reviewed international journal.

Silvestre, I., L. Jordão L., and M. J. Borrego

Personal contribution

IS contributed to the design of the manuscript and wrote the paper.

Abstract

Streptococcus agalactiae remains an important etiological agent of several infectious diseases including neonatal septicemia, pneumonia and meningitis. Although the transition from colonization to invasive infection is not fully understood, the first step in *S. agalactiae* disease is adhesion to host cell surface. *S. agalactiae* expresses a variety of surface-associated and a variety of secreted proteins. Molecular evolution and new genotypic tools have provided insights into the population structure and epidemiology of human *S. agalactiae*; it became clear that *S. agalactiae* strains of certain clonal complexes possess a higher potential to cause invasive disease. Serotype III/ST17 emerged as a hypervirulent clone mostly associated with meningitis in late-onset disease. DNase production and biofilm formation were proposed to explain the leading role of *S. agalactiae* III/ST17 clone. This clone is characterized by the presence of specific gene sets (Srr2, HvgA, Rib, PI-1/PI-2b) and stability among neonatal invasive disease, suggesting that it is well adapted to host environment. However, the emergence of a highly resistance sublineage within *S. agalactiae* III/ST17 clone highlights the role of horizontal genetic transfer in *S. agalactiae* genome evolution. The emergence of serotype IV with the acquisition of the HvgA adhesin, as result of capsular switching in ST17 strains, is also a cause of concern. Consequently, a genotypic surveillance of *S. agalactiae* strains and the development of novel preventive strategies are crucial. Here, we propose to revise *S. agalactiae* genome characteristics and the main virulence factors that contribute to the epidemiological success of clone III/ST17.

Keywords: *Streptococcus agalactiae*; pathogenesis; virulence factors; biofilms; DNases; clone III/ST17; molecular evolution

2.1. Introduction

Streptococcus agalactiae is a common inhabitant of the healthy human gut and urogenital tract, but also an opportunistic pathogen in animals and humans, being the leading cause of human neonatal pneumonia, sepsis, and meningitis [26]. An estimated 2.7 million infants die annually due to perinatal health complications, including at least 90,000 due to invasive *S. agalactiae* disease [27, 28]. This pathogenicity is due to an arsenal of virulence factors that promote colonization, persistence, and onset of invasive infections [37].

S. agalactiae infections usually occur within the first week of life, being designated as EOD. Most infections occur during delivery through aspiration of infected amniotic fluid or direct contact with the colonized vaginal mucosa of the mother, which translates into pneumonia, sepsis and, less commonly, meningitis [4, 22, 29-31]. LOD develops in infants aged >1 week, with most infections becoming evident during the first 3 months of life [22, 32]. After delivery, transmission of *S. agalactiae* to the neonate may occur through horizontal transmission by maternal, nosocomial, or community sources, giving rise to LOD, characterized by bacteremia and meningitis in a high percentage of cases [4, 22, 24, 33, 34].

The precise and unequivocal characterization of *S. agalactiae* strains would be a valuable tool to evaluate streptococcal infections. To achieve this, the method of MLST seems to be discriminatory enough for epidemiological studies, being the standard method to portray *S. agalactiae* clonal population structure providing high intra- and interlaboratory reproducibility [92, 99, 122]. In fact, MLST has been used to determine *S. agalactiae* evolutionary relationships and genetic profiles from different geographic areas, clinical origins, hosts and outbreaks. A ST is defined based on the sequencing and analysis of the internal portion of seven selected *S. agalactiae* housekeeping genes (*adhP, pheS, atr, glnA, sdhA, glcK, tkt*) from the core genome [99]. The combination of seven *loci* alleles assigns a specific single allelic profile or ST. The criterion to group different STs in a CC is based on the same number of alleles shared between the several STs [99, 101]. Until now, 1650 STs have been defined for *S. agalactiae* (https://pubmlst.org/sagalactiae/, accessed at 24-09-2020). Strains belonging to ST17 are considered to account for more than 80% of *S. agalactiae* LOD cases, and often, but not exclusively, are associated with meningitis, supporting the idea that this clone has a particular ability to invade the central nervous system of neonates [37]. As such, ST17 strains should exhibit pathogenic dissimilarities in relation to other genetic lineages, namely ST19, which have been associated with colonization [88].

Here, we propose to revise the main virulence factors and the genomic characteristics of *S. agalactiae* that contribute to the epidemiological success of some clones, particularly regarding the major pathogenic clone III/ST17.

2.2. *S. agalactiae* hypervirulent clone III/ST17

Genome sequencing has provided new insights into the population structure and epidemiology of *S. agalactiae* in humans. The concept of *open pan-genome* in *S. agalactiae* was established since Tettelin *et al.* [112] identified several new specific genes per genome sequenced, reflecting their genomic diversity. In fact, it has been hypothesized that the emergence of *S. agalactiae* hypervirulent clones can be a result of horizontal gene transfer events, which play a significant role in genome evolution of *S. agalactiae* [11, 68]. In the case of mastitis, the exchange of genetic material between the different species that infect the udder (e.g. *S. uberis*, *S. dysgalactiae* ssp. *dysgalactiae* and *Staphylococcus aureus*) might have result in the incorporation of virulence factors (namely inducing antibiotic resistance) into the genome and facilitated *S. agalactiae* adaptation to the host [111, 113]. As such, Gori *et al.* [107] recently hypothesized that *S. agalactiae* strains infecting humans may have emerged from an original ancestral strain infecting animals before crossing to humans.

Decades of investigation have uncovered a remarkable association of *S. agalactiae* hypervirulent clone III/ST17 with an increased ability to cause neonatal invasive disease, particularly meningitis, among LOD cases [32, 68, 123, 124]. It was hypothesized that the acquisition of the integrative and conjugative elements (ICEs) Tn916 and Tn5801, harboring the tetracycline resistance determinant *tetM*, led to the expansion of *S. agalactiae* clones causing human disease, namely CC17 clones, contributing to the increase of neonatal infections [68, 69, 123]. However, how the evolution of hypervirulent ST17 strains provided their unique ability to colonize and infect the human host remains unknown [123]. The stability and dominance of clone III/ST17 among neonatal invasive infections in the past decades indicates an extremely good adaptation to its niche [32]. Indeed, a limited variation in the core genome of CC17 strains (due to recombination) has been reported [68, 69].

The hypervirulent clone III/ST17 has been the focus of genetic and functional studies that identified some of its unique virulence traits, which will be summarized under the following headings. Although it is likely that more than one virulence determinant contributes to the hypervirulence of this ST, the virulence arsenal of ST17 strains has been partially deciphered, linking the β-H/C, serine-rich repeat (Srr) proteins, hypervirulent adhesin (HvgA), protein Rib and pili, with hypervirulence [34, 107, 125]. Accordingly, strains of this genetic lineage are characterized by the presence of specific gene sets, such as Srr2, HvgA, Rib, PI-1/PI-2b [32, 68].

Correlation of *S. agalactiae* invasiveness with DNase activity and biofilm formation have also been proposed to explain the leading role of ST17 strains in neonatal meningitis, as capsular type III/ST17 isolates have been shown to be the main biofilm producers among *S. agalactiae* strains [126-128].

2.2.1. Capsular polysaccharide, CPS

CPS is one of the most important virulence factors involved in bacteria persistence and survival within the host [37, 106]. *S. agalactiae* cell wall is composed of a peptidoglycan-rich layer that serves as a molecular scaffold for anchoring type-specific CPS and surface proteins [110]. CPS mimicry of the host carbohydrate epitopes has been pointed as a facilitator for evasion from the host immune system [37, 106]. During infection, *S. agalactiae* differential infectious capabilities (control of CPS amount, adherence and invasion fitness), should be conferred by up- and downregulation of CPS [106, 129]; however, the understanding of gene regulation by the *cps* operon is limited. CPS is a primary antigen for host recognition, that interfers with the complement-dependent defense pathways, diminishes the phagocytic function of neutrophils, facilitates bacterial internalization and intracellular survival and mediate biofilm formation of *S. agalactiae* in the presence of human plasma [37, 106, 110, 130]. The structure of the CPS of *S. agalactiae* is determined by the *cps locus* genes that encode enzymes responsible for the synthesis of polysaccharides, allowing the classification of *S. agalactiae* strains into capsular polysaccharide-based serotypes Ia, Ib and II to IX [9, 32, 77, 78]. *S. agalactiae* capsular-type III is considered the dominant invasive clone worldwide, being responsible for most cases of neonatal meningitis [36, 37, 44, 85, 87, 89, 90], with a significant percentage of EOD cases, and the majority of LOD [11].

Capsular switching is an important recombination event that occurs through the conjugative transfer of large genomic regions, and has a major role in the diversification of *S. agalactiae* lineages [69]. Capsular type/ST associations have been detected for both colonizing and invasive *S. agalactiae* strains, for example: Ia/ST23; III/ST17; III/ST19; IV/ST196; V/ST1 [69, 131]. The invasive characteristics of capsular type IV strains may result from capsular switching in ST17 strains accompanied by the acquisition of the HvgA adhesion factor that was primarily isolated from hypervirulent ST17 capsular type III strains [37, 87, 90, 132]. Interestingly, capsular type IV might possess relevant genetic features that increase the probability of its involvement in genetic recombination events (supported by the detection of capsular type IV strains with different genetic backgrounds), and this characteristic has been related to an increase of this capsular type [37, 90].

An ongoing emergence of novel virulent *S. agalactiae* clones justifies a continuous epidemiological surveillance [37]. If a putative CPS-based vaccine is developed, this could exert selective pressure over virulent genotypes to switch capsules and escape from vaccine coverage, compromising the success of CPS-based *S. agalactiae* vaccines [90].

2.2.2. β-hemolysin/cytolysin, β-H/C

Hemolytic activity of *S. agalactiae* is due to the ornithine rhamnolipid pigment which is produced by the genes of the *cyl* operon [133, 134, 135]. β-H/C was shown to contribute to *S. agalactiae* colonization and penetration of the chorioamniotic membranes of human placenta by inducing loss of the barrier function in human amniotic epithelial cells [106, 133, 134]. In 2013, hyper-hemolytic *S. agalactiae* strains were isolated from either the amniotic fluid or chorioamniotic membranes of women in preterm labor; as such, hyper-hemolytic strains seem to be proficient in disrupting the amniotic barrier and in penetrating placental membranes [133]. Hemolytic activity of *S. agalactiae* also induces a strong immune response, namely through the formation of Neutrophils Extracellular Traps (NETs) in chorioamniotic membranes (nonhuman primates/murine models of colonization), which play an important role in pathogenesis [133, 134, 136]. NETs consist of nuclear or mitochondrial DNA, as a backbone, with embedded antimicrobial peptides, histones and cell-specific proteases, providing an extracellular matrix to entrap and kill microbes when released by activated neutrophils at sites of infection [136-138]. *S. agalactiae* strains seem resistant to the antimicrobial activity of NETs, likely through increased pigment-mediated antioxidant activity [134, 136]. Thus, hemolysin appear as an important *S. agalactiae* virulence factor, and strains lacking pigment and hemolysin production are often thought to be virulence attenuated [9, 139]. However, there are reports contradicting this hypothesis; e.g. 1) a nonhemolytic hypervirulent *S. agalactiae* strain isolated from a septic neonate [140], and 2) *S. agalactiae* strains of the hypervirulent clone III/ST17 exhibiting very low levels of hemolysis [9, 141]. A possible explanation for these observations would be the lack of correlation between β-H/C production *in vitro* and *in vivo* [9].

2.2.3. Fibrinogen-binding proteins, Fbs

Five fibrinogen-binding proteins (Fbs) have been characterized in *S. agalactiae*: FbsA, FbsB, FbsC (or BsaB), Srr1 and Srr2. Fbs mediate the interaction of *S. agalactiae* with fibrinogen, an abundant protein present in human blood [142]. In general, invasive *S. agalactiae* isolates seem to display stronger fibrinogen-binding affinity in comparison to colonizing strains [37]. FbsA and FbsC promote *S. agalactiae* vaginal colonization due to their ability to adhere to human epithelial cells. FbsC mediates biofilm formation, as its binding to other commensal microbes in the gastrointestinal and/or vaginal tract could facilitate vaginal colonization. FbsC might also mediate *S. agalactiae* dissemination as deletion of FbsC was shown to induce a drastic reduction in strain ability of adherence, invasion of human brain endothelial cells and biofilm formation [37, 143]. FbsB is considered important for *S. agalactiae* dissemination by promoting bacterial invasion of

human epithelial cells [37, 106, 108]. Although FbsA and FbsB are both expressed in the *S. agalactiae* III/ST17 clone, the fibrinogen-binding abilities of the hypervirulent ST17 clone are mainly attributable to FbsB [37, 144]. *S. agalactiae* clone III/ST17 lacks FbsC [37, 68, 123, 125, 145], as *fbsC* gene is not adequately expressed due to a lineage-dependent frameshift mutation [37]. Srr1 and Srr2 were related to *S. agalactiae* adaptation and host specificity [68], as the bacterium expresses either Srr1 or Srr2 on its surface [125]. Srr1 binds to host ECM, stimulating the disruption of the blood brain barrier (BBB), and thereby enhancing meningitis. Binding of Srr1 to fibrinogen seems to be vital for the bacterium adherence to the endothelium of brain vessels, which should occur prior to the development of meningitis [125, 146]. Srr2 promotes *S. agalactiae* attachment to endothelial cells and evidence great affinity for fibrinogen, which may contribute to the increased virulence associated with Srr2-expressing strains [125, 145]. Srr2 is exclusively detected and highly expressed in ST17, promoting virulence, whereas Srr1 is absent in strains from this ST [37, 125].

2.2.4. Hypervirulent adhesin, HvgA

HvgA, a cell-wall anchored surface protein, is a critical virulence trait of *S. agalactiae* in neonatal invasive infection. *S. agalactiae* strains that express HvgA adhere more efficiently to several cell lines, including intestinal epithelial cells and microvascular endothelial cells of the BBB. Thus, HvgA seems to contribute not only to colonization but, more importantly, to the invasion characteristics of hypervirulent clones. Accordingly, all *S. agalactiae* strains of the hypervirulent clone III/ST17 possess HvgA, which is exclusively expressed by CC17 strains. HvgA should confer an enhanced ability for strains to attach to endothelial and epithelial cell lines and thus, to cross BBB in neonates [37, 108, 145, 147]. However, the host receptor for HvgA remains to be elucidated [37, 108, 125, 145, 147].

2.2.5. Alpha-like family surface proteins, Alp

Protein Rib is one of the members of the Alpha-like family surface proteins (Alp) together with Alpha C, Epsilon (Alp1), Alp2, Alp3 and Alp4. These six proteins are encoded by mosaic genes, for which recombination occurs at the same chromosomal *locus*, and are characterized by long *tandem* repeat sequences, whose number of repetitions should have implications for recognition by the host immune system. In fact, it was demonstrated that, during the course of infection, the number of repetitions might suffer internal changes, by deletion events, changing the interaction with the host, and thus mediating invasion [148, 149]. Alp genes have been proposed to belong to horizontally transferred genomic islands. Inter-strain transfer of Alp genes

or gene domains in combination with homologous recombination could explain the broad distribution of these genes among *S. agalactiae* strains, and justify the structural characteristics of the Alp family in terms of mosaicism and immunological properties [150-152]. Nevertheless, the high similarity between the sequences of the different Alp could substantiate the immunological cross-reactivity between different Alp, providing cross-protective immunity, which could be of importance for the development of a vaccine [152].

Regarding ST17 strains, these only carry the Rib protein. Rib should contribute to the virulence potential of ST17 strains, but it also induces protective immunity [123]. A reduction in Rib expression or strains with a smaller Rib were proposed to be more fitted to escape from maternal antibodies, representing a selective advantage to cause neonatal disease [123, 153].

2.2.6. Pili

Pili, cell-wall anchored long filamentous structures extending from the bacterial surface, have been described for several Gram-positive microorganisms [37, 108]. Three genetic pilus islands (PI), PI-1, PI-2a, and PI-2b have been identified in *S. agalactiae*, that encode distinct pilus structures, which mediate specific interactions with host cells [37, 68]; at least one variant or a combination of two PI seem to be present in all *S. agalactiae* isolates [37]. PI consist of five genes that encode for pilus assembly: pilus backbone protein, two ancillary proteins, and two sortase proteins [154]. PI-1 pili do not appear to contribute to bacterial adhesion to the lung, vaginal or cervical epithelial cells but were found to play an important role in evasion from host innate immunity, diminishing by half *S. agalactiae* killing by macrophages [37, 155]. PI-2a pili seem to be specifically involved in adhesion and biofilm formation [37, 156, 157]. PI-2b protein was demonstrated to increase the intracellular survival in macrophages, and has been suggested to promote strain invasiveness and interaction with the host cell, displaying an important role on the infectious process, namely on the BBB penetration [37, 158]. PI-2b pilus variants are almost exclusively present in *S. agalactiae* bovine isolates, as bovine strains mostly lack PI-1 [37, 68]. A strong association between PI and capsular type has been reported [159]. Human *S. agalactiae* strains causing invasive neonatal disease, including all III/ST17, harbor PI-1 plus PI-2b [32, 37, 68, 154]; however, the loss of PI-1 was described in human strains of the CC17 [32, 68, 160, 161]. Martins *et al.* [32] identified, for the first time in Europe, an emerging sublineage characterized by the loss of PI-1 (CC17/PI-2b), simultaneously resistant to macrolides, lincosamides and tetracycline, and exhibiting high-level resistance to streptomycin and kanamycin, which could preclude treatment in people allergic to penicillin. Previously, Teatero *et al.* [68] described the acquisition of MGEs encoding resistance to tetracycline, macrolides and other antibiotics in CC17 strains without PI-1. The loss of PI and the acquisition of MGEs carrying antibiotic resistance

genes contribute to variation among the ST17 *S. agalactiae* lineage [68]. The fact that these MGEs integrated at the same location expected for PI-1, may indicate that the loss of PI-1 and the acquisition of the MGEs occurred simultaneously [68]. Considering the emergence of a highly resistant sublineage within the ST17 hypervirulent clone, it was hypothesized that these highly resistant isolates might be expanding unnoticed [32]. PI-1 has been considered a potential vaccine candidate; as such, the loss or the replacement of this locus may have negative implications on the putative protection offered by pilus subunit-based vaccines [68].

2.2.7. Extracellular DNases

Extracellular DNases have been identified in several species [138]. The genetic background for the observed DNase activity in *S. agalactiae*, first reported in 1980 [162], was established in 2013 with the identification of nuclease A (NucA) gene [163]. Whether other DNases exist in *S. agalactiae* remains to be disclosed. The *nucA* gene of *S. agalactiae* appears unrelated to any other streptococcal DNase coding gene. Its gene product has confirmed DNase activity and seems required for developing a full infection, as decreased mortality was observed in mice inoculated with a mutant strain of reduced nuclease activity [138, 163]. The function of extracellular DNases during pathogenesis is still relatively unclear, and only recently the potential for DNases to be virulence factors has been proposed. One hypothesis would be that scavenging nucleic acids by the action of DNases could provide nutrients during infection or colonization [138]. Another hypothesis would be that streptococcal DNases may be able to effectively eliminate biofilms, formed by other bacteria or regulate the formation of their own, whenever nucleic acids are the fundamental component of a microbial biofilm [137, 138]. A third hypothesis is that DNases could be involved in the bacterium evasion strategy to prevent neutrophil activation and contribute to NETs degradation, thereby promoting resistance to NETs-mediated antimicrobial activity of neutrophils and contributing to pathogenesis [137, 138]. Accordingly, part of the success of the clone III/ST17 could rely on the activity of extracellular DNases on the NETs released by neutrophils [128, 164, 165]. In fact, the invasive disease caused by another streptococcal species, *S. pyogenes*, has reappeared due to the emergence and global dissemination of the highly virulent clone M1T1, which exhibits two unique bacteriophages (acquired through horizontal transfer) that encode the potent DNase Sda1/SdaD2 and the superantigen SpeA [166, 167]. DNases are also associated with MGEs, and can thus be transmitted between strains and even species [138]. The hypothesis that *S. agalactiae* can acquire DNases through MGEs remains to be explored.

2.2.8. Biofilm formation

The adherence capacity of *S. agalactiae* to host cells facilitates bacterial cell aggregation and biofilm formation, an important virulence determinant that has been linked to colonization and progression to disease. Biofilms are sessile colonies embedded in a self-produced matrix of extracellular polymeric substances (EPS), offering protection in severe environmental conditions that can include antimicrobials, extreme pH, and immune cells, promoting the maintenance of a bacterial population, and thus contributing to chronic infection and heavy colonization [1, 168-170]. *S. agalactiae* biofilm formation seems to be enhanced by CPS [37, 107, 130, 156], and the deletion of the gene encoding biofilm regulatory protein A (BrpA) was shown to impair both biofilm formation and the ability of the bacterium to colonize and invade [107, 110]. The environmental factors that favor biofilm formation by *S. agalactiae* strains are contradictory. In fact, results from studies addressing the biofilm-forming capacities of strains from different origin (colonization or infection), or with different genotypic features (capsular type or ST), and the correlation between biofilm formation and glucose concentration or pH value, seem to vary [126, 127, 168, 169, 171-174]. A correlation of *S. agalactiae* III/ST17 strains invasive capability with biofilm formation was also proposed [126, 127]. As such, several *in vitro* studies evaluated biofilm formation by different *S. agalactiae* strains, namely within CC17, providing conflicting data; i.e. CC17 strains were described as strong [126, 169] but also as weak biofilm producers [170], probably due to methodological differences. The discrepancies among studies regarding the ability of CC17 strains to form biofilms *in vitro* can be also explained by the difference in PI distribution [169]. Variability in both presence and sequence of pilus loci has been investigated, and an inhibitory effect of PI-1 in biofilm formation in CC17 and CC19 strains was observed [170]. A strong biofilm phenotype in CC17 strains carrying exclusively PI-2b was also described; nonetheless, no precise association between PI and biofilm production could be established [169].

2.3. Conclusion

Despite the strategies implemented for the prevention of *S. agalactiae*, this pathogen remains the leading cause of neonatal invasive bacterial infection. *S. agalactiae* reflects a unique adaptation to host immune pressure through genome constant modification. Horizontal gene transfer should play a significant role in *S. agalactiae* genome evolution, driving to clonal expansion of the best-fitted strains. Despite the low levels of ST17 core genome recombination, the acquisition of MGEs encoding antimicrobial resistance and the loss of PI-1 surely contribute to the diversity among *S. agalactiae* strains. The phenomena involved in *S. agalactiae* host colonization and pathogenesis, and the emergence of new hypervirulent clones, like clone

III/ST17 sublineages, may have significant implications for the prevention and management of *S. agalactiae* disease. The understanding of the complex, and yet to be fully unveiled, host-*S. agalactiae* interaction is required to identify individuals at high risk of invasive disease and for engendering effective preventive and curative therapies.

Chapter III

Genomic insights on DNase production in *Streptococcus agalactiae* ST17 and ST19 strains

The contents of this chapter were submitted (with minor changes) to a peer reviewed international journal and were published as a preprint (DOI: 10.1101/2020.12.09.418327).

Silvestre, I., A. Nunes, V. Borges, J. Isidro, C. Silva, L. Vieira, M. J. Borrego, and J. P. Gomes.

Upon the preparation of this manuscript, additional data concerning the ST17 and ST19 strains under study were obtained. This was the case of putative virulence determinants which revealed ST-specific particularities (Supplementary Table S3.4, Supplementary Figures S3.5). The antimicrobial susceptibility phenotype regarding penicillin G, vancomycin, trimethroprim-sulfamethoxazole, erythromycin and clindamycin was also determined, as well as the molecular mechanisms underlying the resistance phenotype (Supplementary Table S3.5, Supplementary Figures S3.6).

Personal contribution

IS performed the experimental procedures, co-analyzed the data, and co-wrote the paper.

Abstract

Streptococcus agalactiae evasion from the human defense mechanisms has been linked to the production of DNases. These were proposed to contribute to the hypervirulence of *S. agalactiae* ST17/capsular-type III strains, mostly associated with neonatal meningitis. We performed a comparative genomic analysis between ST17 and ST19 human strains with different cell tropism and distinct DNase production phenotypes. All *S. agalactiae* ST17 strains, with the exception of 2211-04, were found to display DNase activity, while the opposite scenario was observed for ST19, where 1203-05 was the only DNase(+) strain. The analysis of the genetic variability of the seven genes putatively encoding secreted DNases in *S. agalactiae* revealed an exclusive amino acid change in the predicted signal peptide of GBS0661 (NucA) of the ST17 DNase(-), and an exclusive amino acid change alteration in GBS0609 of the ST19 DNase(+) strain. Further core-genome analysis identified some specificities (SNVs or indels) differentiating the DNase(-) ST17 2211-04 and the DNase(+) ST19 1203-05 from the remaining strains of each ST. The pan-genomic analysis evidenced an intact phage without homology in *S. agalactiae* and a transposon homologous to TnGBS2.3 in ST17 DNase(-) 2211-04; the transposon was also found in one ST17 DNase(+) strain, yet with a different site of insertion. A group of nine accessory genes were identified among all ST17 DNase(+) strains, including the Eco47II family restriction endonuclease and the C-5 cytosine-specific DNA methylase. None of these loci was found in any DNase(-) strain, which may suggest that these proteins might contribute to the lack of DNase activity. In summary, we provide novel insights on the genetic diversity between DNase(+) and DNase(-) strains, and identified genetic traits, namely specific mutations affecting predicted DNases (NucA and GBS0609) and differences in the accessory genome, that need further investigation as they may justify distinct DNase-related virulence phenotypes in *S. agalactiae*.

Keywords: *Streptococcus agalactiae*; Whole Genome Sequencing; DNases; ST17; ST19

3.1. Introduction

Streptococcus agalactiae is a major human pathogen which, although present as a commensal bacteria in the gastrointestinal and genitourinary tract, is a leading cause of neonatal morbidity and mortality worldwide, often associated with neonatal meningitis, pneumonia and sepsis [123, 138]. The ability of *S. agalactiae* to cause a wide range of disease may be due to an extensive arsenal of virulence factors that contribute to pathogenicity [138, 175], which includes biofilm formation and the production of DNases [138, 163, 176]. In fact, DNases would be able to destroy neutrophil extracellular traps (NETs) released by neutrophils upon microbial infection, when they are recruited to the infectious site to restrain bacterial spreading [164, 165, 176, 177], as NETs are composed of DNA, histones, proteolytic enzymes and other peptides. However, many pathogens have evolved mechanisms to evade NET-mediated entrapment and killing in what appears to be a widespread strategy to allow pathogen proliferation and dissemination [138, 176, 177]. DNases can also reduce Toll-like receptor 9 (TLR-9) signaling to dampen the immune response and produce cytotoxic deoxyadenosine to limit phagocytosis [138, 176]. Streptococcal DNases may be further involved in interacting with other microbial communities through bacterial killing and disruption of competitive biofilms, or be able to control their own biofilm production. In addition, the production of DNases may provide nutrients during colonization and infection by scavenging nucleic acids of neutrophils, as NET formation ultimately leads to cell death. This may explain why many streptococci possess both secreted and cell-wall-anchored DNases. However, the role of DNases during pathogenesis is not completely understood and may be wide ranging, extending beyond the direct interference with the host immune response [138, 177].

High nuclease activity by *S. agalactiae* strains was firstly reported in 1980 [162]. Advances in molecular biology, namely the use of WGS, evidenced homology between the DNases of different streptococcal species [138]. In 2013, Derré-Bobillot and co-authors [163] identified seven genes putatively encoding secreted DNases in *S. agalactiae* NEM316 (*gbs0153, gbs0382, gbs0609, gbs0661, gbs0712, gbs0885* and *gbs0997*), including the major nuclease (NucA) coding gene (*gbs0661*). Their results showed that the loss of NucA, which is secreted and capable of degrading NETs, resulted in reduced *S. agalactiae* NEM316 virulence [163, 176]. Besides, it seems likely that other DNases exist in *S. agalactiae* but have yet to be identified [138]; as such, recent genome screenings confirm the presence of potential extracellular DNases in mostly all available *S. agalactiae* genome sequences [176].

S. agalactiae capsular-type distribution and predominance could change over time; however, it became clear that strains of certain CCs possess a higher potential to cause invasive

disease, while other harbor mainly colonizing strains [37]. *S. agalactiae* hypervirulent ST17/capsular-type III strains gained special interest due to their strong association with neonatal meningitis and LOD of the newborn [37, 87, 90, 99, 103, 124, 147, 159, 178]. These strains display secreted and surface-exposed factors, including DNases [88, 124, 153]; accordingly, ST17 strains were recently reported as DNase producers while the DNase non-producer phenotype was only found among CC19 strains, which are well adapted to the vaginal mucosa, evidence a poor invasion ability, and were less often related to meningitis [37, 128].

The aim of the present study was to shed some light on the genetic background of DNase production in *S. agalactiae* strains from different clinical scenarios. For this purpose, we performed a comparative genomic analysis between ST17 and ST19 human strains with different cell tropism (invasive *vs* carriage) and diverse DNase production phenotypes.

3.2. Materials and methods

3.2.1. Strain collection

In this study, a total of 33 clinical *S. agalactiae* strains of human origin (28 colonizing and 5 invasive) were analyzed, including isolates with differential DNase activity determined through prior assays [128]. The selected strains belong to the collections of the Portuguese National Institute of Health (Instituto Nacional de Saúde Doutor Ricardo Jorge, INSA, IP) and the Institute of Medical Microbiology and Hygiene (IMMH), Ulm University, Germany.

For comparative purposes, *S. agalactiae* reference strains exhibiting distinct DNase activities were used: NEM316 (genotype: III/ ST23) (ATCC 12403, GenBank accession number NC004368); COH1 (genotype: III/ST17) (GenBank accession number AAJR01000000); 2603V/R (genotype: V/ST110) (ATCC BAA-611, GenBank accession number NC004116).

3.2.2. Bacterial culture

Bacteria were grown in Columbia agar supplemented with 5% sheep blood (Biomérieux, Marcy l'Etoile, France) at 5% CO_2 for 24 h. These cultures were used to proceed to DNase activity assays, antimicrobial susceptibility testing, and to inoculate cultures on fresh Todd Hewitt broth (THB) (BD®, New Jersey, USA) supplemented with 0.5% yeast extract, which were allowed to incubate without shaking at 37ºC with 5% CO_2. Cell growth was monitored by measuring the optical density (OD) at 600 nm (OD_{600}). At middle exponential phase (OD_{600}=0.2-0.5), 1 ml of bacterial cells was collected by centrifugation (3000 rpm, 10 min), resuspended in 200µl of PBS and immediately stored at -20ºC, before further DNA extraction.

3.2.3. DNase activity, qualitative assays

DNase production phenotype was assessed qualitatively by growing *S. agalactiae* strains on DNA-methyl green agar plates (Oxoid, Basingstoke, England), that allowed the identification of DNase production. The results were interpreted after 24 h of incubation at 37ºC in an atmosphere of 5% CO_2. Strains were considered DNase producers (DNase(+)) when displaying transparent halos around colonies of *S. agalactiae* in DNA-methyl green agar plates. *S. agalactiae* strains COH1 and NEM316 were used as positive controls and 2603V/R as a negative control (DNase(-)).

3.2.4. DNase activity, semi-quantitative and quantitative assays

In order to confirm the results of the qualitative tests, which pose challenges for interpretation based on visualization, a semi-quantitative assessment of the activity of the DNases was further performed according to the described by Sumby and co-authors [179], with modifications. Briefly, *S. agalactiae* strains were inoculated in 5 ml of THB supplemented with 0.5% yeast extract, culture supernatant from stationary phase was centrifuged (10 min., 3000 rpm), syringe filtrated (0.2 µm), before incubation at 37°C with 1 µg of double stranded DNA (*atr* amplicon [99]) in the presence of 1x M buffer (Roche). Filtrated supernatants (10 µl) were incubated for 1h, 2h, 4h and "overnight" (~ 17h), in order to evaluate the digestion of DNA in a final volume of 50 µl. Nuclease reaction was stopped by adding EDTA (0.5 M, pH 8.0) at 4ºC. DNA digestion was analyzed by electrophoresis in 1% agarose gel. A negative control consisting on a reaction mixture without supernatant was used in all experiments.

In order to confirm the results of production / non-production obtained by the methods previously described, a quantitative evaluation of the activity of DNases was further performed for a selection of *S. agalactiae* strains. The quantitative DNase assays were performed as described above and the amount of dsDNA present in each sample was determined by measuring at the same time points of incubation, using Quant-iT™ PicoGreen™ dsDNA assay kit (Thermo Fisher Scientific, USA), according to manufacturer's instructions. Briefly, 1 ml of 1x Quant-iT PicoGreen was added to an equal volume of each supernatant, previously diluted in 1x TE buffer. After 5 minutes of incubation at room temperature (RT), in the dark, we proceeded to the fluorescence measurement (Fluorimeter – Anthos Zenith 3100) in 96 well microtiter plates (Corning 96 Well Clear Flat Bottom Polystyrene Black TC, Costar, USA). To calculate the concentration of the remnant DNA in each sample, a standard curve with four solutions of phage Lambda DNA of known concentrations (1, 10, 100 and 1000 ng/ml) was used. Each obtained fluorescence value was subtracted of the blank solution (PicoGreen 1x + 1x TE buffer at a ratio

of 1:1) fluorescence value. *S. agalactiae* strains COH1 and NEM316 were used as positive controls and 2603V/R as a negative control.

3.2.5. Nucleic acid isolation

Genomic DNA was extracted as previously described [180], with minor changes. Briefly, bacterial cells were subjected to a high-speed centrifugation (14000 rpm) for 10 min at 4ºC. The pellet was digested for 2 h at 37ºC with 200 µl of Tris-EDTA buffer, pH 8.0, containing 10 U mutanolysin (Sigma-Aldrich, St. Louis, USA) and 15 mg/ml lysozyme (Sigma-Aldrich, St. Louis, USA) before treatment with 10 mg/ml proteinase K (Roche, Penzberg, Germany). Subsequently, DNA was extracted using the NucliSENS® EasyMag® (BioMérieux, Marcy l'Etoile, France) or the Isolate II Genomic DNA kit (Bioline, Tennessee, USA), according to the manufacturer's instructions. DNA quality, purity and concentration were evaluated by agarose gel electrophoresis, spectrophotometry (NanoDrop, ThermoFisher Scientific, Massachusetts, USA) and fluorimetry (QubitTM, ThermoFisher Scientific, Massachusetts, USA), respectively. Then, DNA samples were sent to sequencing in Innovation and Technology Unit, at the Portuguese National Institute of Health.

3.2.6. Whole genome sequencing, de novo assembly and annotation

Previously selected *S. agalactiae* ST17 and ST19 strains (n=33) were all subjected to WGS, as previously described [181]. Briefly, high-quality DNA samples were subjected to dual-indexed Nextera XT Illumina library preparation (Illumina, USA), prior to cluster generation and paired-end short-read high throughput sequencing (2x150bp or 2x250bp) on an Illumina MiSeq or NextSeq550 equipment (Illumina, USA), according to the manufacturer's instructions. Genomes were *de novo* assembled using the INNUca v3.1 pipeline (https://github.com/B-UMMI/INNUca) [182], an integrative bioinformatics pipeline that consists of several integrated modules for reads QA/QC, *de novo* assembly and post-assembly optimization steps. Briefly, after reads' quality analysis using FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and cleaning with Trimmomatic v0.36 (http://www.usadellab.org/cms/?page=trimmomatic) [183], genomes were *de novo* assembled with SPAdes 3.10 (http://bioinf.spbau.ru/spades) [184], and subsequently improved using Pilon v1.18 [185].

Some samples were also subjected to long-read nanopore sequencing (Oxford Nanopore Technologies-ONT) in order to increase the contiguity of genome sequences. Briefly, genomic libraries were prepared using the Rapid Sequencing kit (ONT), according to the manufacturer's

protocol. A MinION device (ONT) USB-connected to a laptop computer with i7 Intel processor and 16 GB RAM, was used to sequence the libraries in two R9.4 type flow cells. The MinKNOW software (v19.06.8; ONT) was used to configure run parameters, data acquisition and real-time base-calling (Guppy v3.0.7; ONT). The quality of the MinION reads was assessed using MinIONQC (https://github.com/roblanf/minion_qc). Canu 1.9 was used for error correction of reads, trimming (removing adapters and breaking chimeras) and assembly with genome size of 2.1 Mbp as input [186]. Curation of Canu assemblies was performed using a parallel mapping strategy: 1) of MinION trimmed reads with Minimap2 (https://github.com/lh3/minimap2) [187], and 2) of Illumina trimmed reads with Snippy v4.6 (https://github.com/tseemann/snippy) software, followed by polishing with Pilon v1.18. The later was ran for multiple iterations until the polished genomes remained unchanged.

Draft genome sequences were annotated with both RAST server v2.0 (http://rast.nmpdr.org/) [188] and Prokka v1.12 [189]. For simplification purposes, NEM316 locus tag (GenBank accession number NC004368.1) was adopted to designate all gene hits identified in the subsequent analyses.

3.2.7. Core-genome single nucleotide variant (SNV)-based analysis

The genetic relatedness among strains was evaluated by a reference-based mapping strategy using Snippy v4.6 software. For each isolate, quality improved Illumina reads were mapped against a representative draft short-read-only assembled genome (2211-04 was used as mapping reference for all 33 strains as well as for ST17 cluster, while 1203-05 was used as mapping reference for ST19 cluster). SNV calling was performed on variant sites that filled the following criteria: *i)* minimum mapping quality of 20; *ii)* minimum number of reads covering the variant position ≥10; and *iii)* minimum proportion of reads differing from the reference of 90%. Core-SNVs were extracted using Snippy's core module (*snippy-core*). All putative SNVs/indels were carefully inspected and confirmed using the Integrative Genomics Viewer (IGV) v2.8.2 (http://software.broadinstitute.org/software/igv/) [190]. MEGA7 software (http://www.megasoftware.net) [191] was applied to calculate matrices of nucleotide distances and perform phylogenetic reconstructions over the obtained core-genome SNV alignment by using the Neighbor-Joining method [192] with the Maximum Composite Likelihood model to compute genetic distances [193] and bootstrapping (1000 replicates) [194].

3.2.8. Extended genetic diversity analysis

The PubMLST online platform (http://pubmlst.org/) was used for *in silico* MLST. For strains for which the *de novo* assembled genome could not be obtained (n=9) (since did not pass FastQC module QA/QC), MLST locus sequences were retrieved by a read-based mapping approach. Alleles and STs not previously described were deposited in the *S. agalactiae* MLST database. CCs were defined using eBurst analysis (http://eburst.mlst.net/v3/mlst_datasets/) [195].

The occurrence and structure/diversity of CRISPR-Cas systems (CRISPR-associated proteins) was predicted for all assembled genomes (n=12 for ST17 and n=12 for ST-19) with both CRISPRCasFinder (http://crisprcas.i2bc.paris-saclay.fr/) and CRISPRone (http://omics.informatics.indiana.edu/CRISPRone/) [196] web tools (both accessed in May 2020). Only intact elements with the highest confidence score level were considered as legitimate hits. The occurrence and structure of each hit CRISPR-Cas system was confirmed by RAST annotation and reads-based mapping.

In order to obtain an overview on the repertoire of putative virulence genes carried by each strain, a vast database was constructed, enrolling: *i)* previously identified virulence determinants [11, 69] and *ii)* genes identified after querying draft genome sequences against the Virulence Factors Database (VFDB; http://www.mgc.ac.cn/VFs/; last updated at 30th September 2020) [197], using the VFanalyzer platform (accessed in October 2020). The presence/absence and/or variability of each putative identified hit was confirmed by an assembly-free strategy using Snippy 4.6, while BLASTp against the non-redundant (nr) protein sequences database was used to assess the potential protein function. The identified hits were then submitted at to PubMLST online platform for allele determination.

As a means to potentiate isolate discrimination, assemblies were also analyzed using PHASTER (https://phaster.ca/) [198] to determine the presence of prophages. Only hits with intact phages were considered for further analysis, given that the detection of phage fragments could result from the assembly process hampering the distinction between presence of intact phages and phage remnants that were excised during the isolate evolutionary process (for which the biological importance is even more uncertain). The identification of other mobile genetic elements (MGEs) was carried out by manual inspection of annotated sequences, and further confirmed through BLASTn and BLASTp searches against the nr and WGS databases. In particular, the presence of plasmids was assessed through PlasmidFinder 2.0 (http://cge.cbs.dtu.dk/services/PlasmidFinder/), using default parameters with both trimmed reads and assemblies.

Finally, global pan-genome analyses were carried out for all assemblies using Roary v3.8.0 software [199] with a BLASTp minimum percentage of identity set to 95% and without

splitting paralogs. In order to obtain referential locus tags, gene hits found to be present/absent exclusively in 'exception strains' for DNase activity, were confirmed through reads' mapping and Blastn search (90% query coverage with 70% identity) against publicly annotated *S. agalactiae* closed or draft genomes.

3.2.9. DNases' structural characterization

For all DNases, the respective predicted protein sequences were searched for the presence of domains and/or motifs indicative of nucleases. SignalP-5.0 Server (https://www.cbs.dtu.dk/services/SignalP/) was used to check the presence of signal peptides. Putative structures were predicted using SWISS-MODEL (http://swissmodel.expasy.org/).

3.2.10. Data availability

All raw sequence reads used in the present study were deposited in the European Nucleotide Archive (ENA) under the study accession number PRJEB41294 (Supplementary Table S3.1).

3.3. Results

3.3.1. DNase activity assays

Of the total 33 clinical *S. agalactiae* selected strains, 15 belonged to ST17 and 18 belonged to ST19. Based on qualitative and semi-quantitative assays, all *S. agalactiae* ST17 strains, with the exception of the clinical strain 2211-04, were found to display DNase activity. The opposite scenario was observed for ST19, where DNase activity was only seen for one (1203-05) out of the eighteen ST19 clinical strains. The presence or absence of DNase activity was further assessed by quantitative assays for a selection of *S. agalactiae* strains from both STs (Figure 3.1). Confirming the previous observations, the ST17 2211-04 clinical strain was the only ST17 strain lacking DNase activity (congruent with the negative control 2603V/R), while ST19 clinical strains (with exception of 1203-05) showed no ability to degrade DNA, whose amount remained nearly constant after 17h. Interestingly, two distinct DNA digestion profiles were observed for the remaining ST17 clinical strains, when compared with the two positive controls, *S. agalactiae* COH1 and NEM316. While some of them mirrored the medium DNA digestion of COH1, others exhibited an almost total DNA digestion (<10% DNA remained after 17h),

reflecting a high production of extracellular DNases like reference NEM316. On the other hand, the ST19 1203-05 clinical strain appeared isolated, revealing only a weak DNase activity.

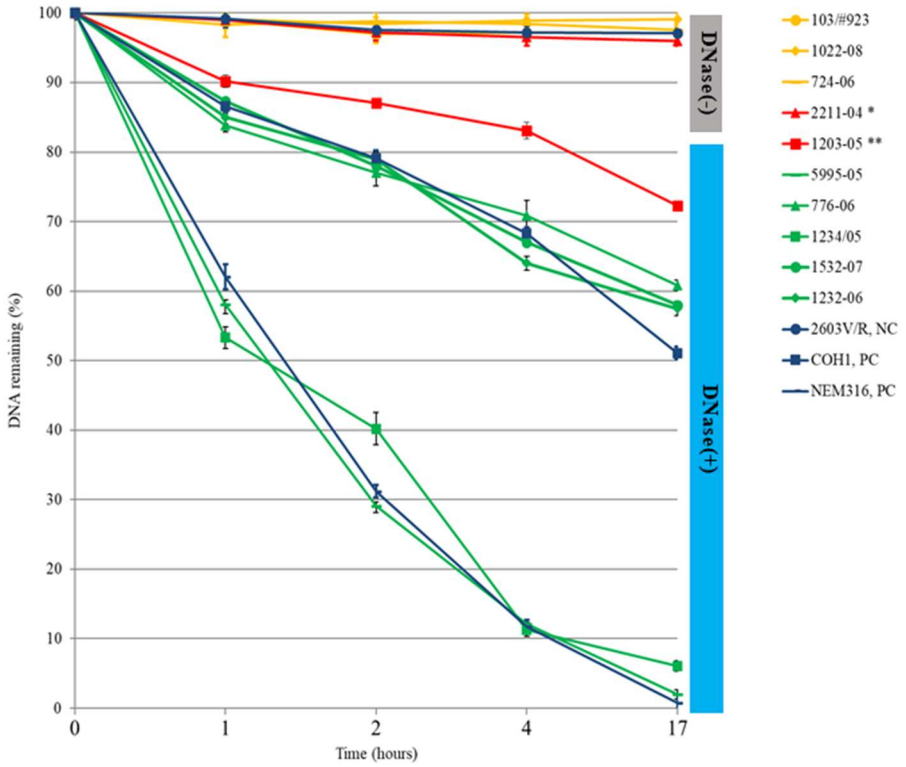


Figure 3.1 Quantitative DNase assays displaying differential DNase activity between *S. agalactiae* strains over time. *S. agalactiae* culture supernatant (10 µl) incubated with 1 µg of DNA (amplicon *atr*) for 1h, 2h, 4h, overnight at 37°C. Fluorescent PicoGreen dye (Invitrogen) was used to quantify the dsDNA. The graphic displays the mean values of the experiment performed in duplicate. Error bars show ± standard deviation. DNase(-) ST19 strains are shown in yellow, while DNase(+) ST17 strains are depicted in green. The two exceptions are displayed in red, with * representing the DNase(-) ST17 strain and ** representing the DNase(+) ST19 producer. All controls are shown in dark blue (PC, Positive Control; NC, Negative Control).

3.3.2. Strains' phylogenetic relationship

Considering the observed DNase activity exceptions within each ST, we first assessed the genetic relationship between 2211-04 and 1203-05 and the remaining clinical strains. We observed that the core-genome SNV-based phylogenetic analysis clearly segregates strains based on each ST (Figure 3.2), resulting in a strong two major-branches tree, where neither ST17 2211-04 nor ST19 1203-05 strains showed up in isolated branches. Indeed, clinical strains from ST17 displayed a mean of 163.2±5.3 core-SNVs among them but had a mean distance of 7342.2±33.2 core-SNVs to ST19 strains, which differ themselves by a mean of 466.5±7.3 core-SNVs (Figure 3.2A). Strains' relatedness within each individual ST was also explored by maximizing the core-

genome under evaluation (Supplementary Figure S3.1). Despite only a slight increase in the number of core-SNVs was observed (mean core-SNVs of 258.9±4.8 within ST17 and mean core-SNVs of 526.2±5.2 within ST19), we were able to identify the clinical strains genetically closest to ST17 2211-04 and ST19 1203-05 strains. For ST17, 2211-04 form an independent phylogenetic clade with 1232-06 (differing by 83.0±9.0 core-SNVs), while for ST19, 1203-05 was found to be genetically closest related to 461-05, 724-06, 1434-05, 1237-07 and 225-06 strains (differing only by a mean of 41.8±5.7 core-SNVs from these strains).

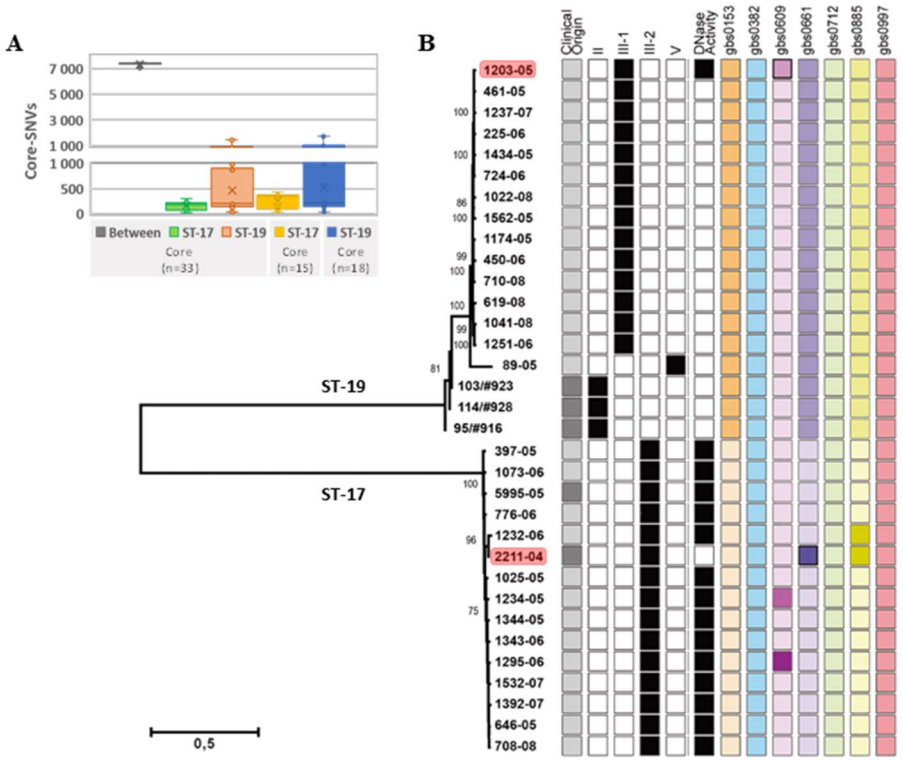


Figure 3.2 DNase-based genotypes and phenotypes. A. Box Plots depicting the number of SNVs observed within the overall population and within each ST. The three left plots are based on a core-genome enrolling all the 33 strains, whereas the two right plots are relative to a core-genome enrolling solely same-ST strains (n=15 for ST17 and n=18 for ST19), in order to maximize the genome shared for each sub-set of strains. The horizontal line within each box marks the median, while the cross ("x") represents the medium value. **B.** The core-genome SNP-based phylogenetic tree was reconstructed using the Neighbor-Joining method [192] with the Maximum Composite Likelihood model [193]. Bootstraps values are shown next to the respective nodes. Colonization strains are represented by grey light squares, while strains from invasive infection are in grey dark squares. For each isolate, the black squares represent the respective capsular-type as well as the presence of DNase activity (illustrated in Figure 3.1). Both the DNase(-) ST17 and the DNase(+) ST19 strains are highlighted in red. For each DNase gene, the different alleles found for each clinical strain are represented by distinct color tones (proximal color tones do not reflect genetic proximity). Genes were designated according to NEM316 genome (GenBank accession number NC004368.1).

3.3.3. Genetic variability of DNase genes

We then proceeded with a more focused analysis where the potential genetic basis behind the observed DNase activity was evaluated. All known loci putatively encoding secreted DNases in *S. agalactiae* were analyzed. While some genes were found to be conserved among all strains (the paralogs *gbs0382*, *gbs0712* and *gbs0997*), or to fully separate strains according to ST (*gbs0153*), specific alleles were exhibited by ST19 1203-05 and ST17 2211-04 clinical strains in *gbs0609* and *gbs0661*, respectively (Figure 3.2B), which may contribute to their exception character regarding DNase activity. For instance, the weak DNase(+) ST19 1203-05 strain was found to harbor an exclusive missense alteration in *gbs0609*, corresponding to the substitution of a negative-charged amino acid by a positive-charged one (Glu261Lys). On the other hand, for NucA coding gene (*gbs0661*), an exclusive non-synonymous alteration (Thr10Ala) affecting the predicted signal peptide was exhibited by the DNase non-producer ST17 2211-04 strain. When compared to the well-studied *gbs0661* of DNase(+) NEM316, all 33 *S. agalactiae* strains additionally displayed one non-synonymous (T296C; Pro99Leu) and two silent (G99A and G228A) substitutions, despite all DNase(-) ST19 strains harbored an extra silent alteration (A423G). Curiously, the replacement of the proline residue by a leucine fall within β-strand 4 of NucA, according to its predicted structure for NEM316 [200], which together with β-strands 9 and 10 constitute the deepest recess of the nuclease active site (Supplementary Figure S3.2). Despite the detected amino acid replacements may hypothetically explain the differences in DNase activity observed in the present study, to our knowledge they have not yet been subjected to any mutagenesis assays, so their role in the enzyme secretion/activity remains unknown.

3.3.4. Core-genome analysis

We then searched for other genetic aspects that could explain the observed DNase activity exceptions. Globally, a total of 35 core-SNVs/indels were found to differentiate DNase(-) ST17 2211-04 from all DNase(+) ST17 strains (Table 3.1). With exception of five core-SNVs falling within non-coding sequences, variants (17 non-synonymous, 11 synonymous, one frameshift deletion and one conservative in-frame insertion) were found to affect genes mostly related to metabolism (n=13) as well as to cellular processes and signaling (n=11), such as those encoding the ATP-dependent Clp protease ATP-binding subunits ClpC and ClpX, the molecular chaperone DnaK, and the class I S-adenosylmethionine (SAM)–dependent methyltransferase. Regarding the core-SNVs affecting non-coding regions, none of them was found to interrupt any annotated regulatory element (like ribosome binding sites and transcriptional terminators, for instance).

Nevertheless, one SNV apparently falls within a region similar to a transposase C-terminal end (pseudogene).

Table 3.1 Core-SNVs/indels between DNase(-) 2211-04 and all DNase(+) ST17 strains.

2211-04 NODE	NODE position (bp)	Change in DNase(+) ST-17 strains	Product	NEM316 old locus tag [a]
1	6712	T518C (Val173Ala)	Uracil phosphoribosyltransferase	gbs1635
2	64811	G1071A (Gly357Gly)	Major cardiolipin synthase ClsA	gbs1088
2	67278	T393C (His131His)	Formate - tetrahydrofolate ligase 1	gbs1089
2	80682	T707C (Phe236Ser)	ABC transporter ATP-binding protein	gbs1103
3	36869	T225C (Val75Val)	Protein RecA	gbs2047
3	77527	T126C (Ile42Ile)	Glycine betaine ABC transporter permease	gbs2089
3	112842	G10A (Ala4Thr)	LysM peptidoglycan-binding domain-containing protein	gbs2107
4	25790	A1548G (Gly516Gly)	Oligoendopeptidase F, plasmid	gbs0779
6	26679	A301G (Asn101Asp)	Hypothetical protein	--- [b]
6	26969	10delA (Ile4fs) [c]	Hypothetical protein	--- [b]
7	4092	G1153A (Val385Ile)	ATP-dependent Clp protease ATP-binding subunit ClpC	gbs1376
7	11190	243-244insAAACGTGCT (Ala81-Leu82insLysArgAla)	ATP-dependent Clp protease ATP-binding subunit ClpX	gbs1383
7	11616	G18T (Leu6Phe)	Hypothetical protein	--- [d]
7	34581	C458T (Ser153Leu)	Putative membrane protein	gbs1406
7	69209	T1018C (Cys340Arg)	Capsule biosynthesis protein CapA	gbs1441
9	42329	A-->G		IGR (gbs0664-gbs0665)
9	**45552**	**G28A (Ala10Thr)**	**Nuclease A**	**gbs0661 [e]**
9	60220	A2054G (His685Arg)	PII-type proteinase	---
11	7603	T789C (Arg263Arg)	Glucose-6-phosphate isomerase	gbs0437
12	12947	T141C (Asp47Asp)	50S ribosomal protein L23	gbs0060
12	50086	T986C (Ile329Thr)	Amidophosphoribosyltransferase	gbs0025
13	35373	T-->C		IGR (gbs1939-gbs1940)
14	6064	A309G (Val103Val)	Glycogen debranching enzyme	gbs1288
14	45399	T118C (Cys40Arg)	Ribose-5-phosphate isomerase A	gbs1256
17	20340	A96G (Ser32Ser)	23S rRNA methyltransferase	gbs1028
18	32253	T156G (Asp52Glu)	Iron-sulfur cluster assembly protein SufB	gbs0141
19	14314	C1220T (Thr407Ile)	1,4-alpha-glucan branching enzyme GlgB	gbs0871
20	37252	T56A (Ile19Asn)	Cobyric acid synthase	gbs0900
21	22836	A193G (Ile65Val)	Class I SAM-dependent methyltransferase	gbs1563
24	197	T-->G	Transposase C-terminal end (truncated) (Y)	gbs0599 (Y) / gbs0620 (Y)
25	19551	T79C (Leu27Leu)	Heme response regulator HssR	gbs2009
26	15787	A234G (Lys78Lys)	Siderophore transport system ATP-binding protein YusV	gbs1462
28	5613	T106A (Ser36Thr)	Chaperone protein DnaK	gbs0096
30	9399	T-->C		IGR (gbs1532-gbs1533)
36	8124	G-->T		IGR (gbs0627-gbs0628)

[a] Open Reading Frames (ORFs) designations according to NEM316 genome (GenBank accession number NC_004368.1)
[b] Corresponds to NEM316 locus tag GBS_RS09875; smaller ORF (1362bp; 454aa)

[c] The DNase(-) GBS-2211-04 possesses an intact ORF of 1533bp (511aa), while the 1-bp deletion harbored by all ST17 DNase(+) strains yields a frameshift change that result in a smaller 1497bp ORF (499aa). In addition, 1232-06 possesses an extra 1-bp deletion that leads to a premature codon stop, resulting in two ORFS (390bp and 981bp) (data not shown)
[d] Corresponds to NEM316 locus tag GBS_RS07280
[e] Gene putatively encoding secreted DNase [163]
Y Pseudogene

On the other hand, for ST19, 24 core-SNVs/indels were found to discriminate the weak DNase(+) 1203-05 from all DNase(-) ST19 strains (Table 3.2). Of the 20 core-SNVs occurring in coding sequences, approximately two thirds originate missense variants, and affect genes encoding hypothetical proteins as well as genes related to bacterial stress response, defense mechanisms or metabolism. From these, we highlight two genes encoding the GTPase Obg and the excinuclease ABC subunit B that are related to bacterial stress response and DNA repair. Interestingly, one core-SNV (G297A) involves the exclusive gain for 1203-05 of a premature stop codon (at 99aa) in the TetR/AcrR family transcriptional regulator gene, likely producing two shorter proteins (99aa and 56aa instead of the original 183aa). Two insertions (of 1bp and 18bp) occurring in genes coding for a nucleoid-associated protein and an hypothetical protein were also found to differentiate all ST19 DNase(-) strains from the weakly DNase(+) 1203-05, yielding proteins with distinct length sizes. Regarding the core-SNVs falling in non-coding sequences, one occurred in the ribosome binding site (RBS) of a *gbs1083* homolog. While all ST19 DNase(-) possess a GGCAGG RBS sequence, the weakly DNase(+) displays an exclusive GGCAGA sequence, which may hypothetically influence the ribosome binding and, consequently, the protein (threonine/serine exporter family protein) expression.

Table 3.2 Core-SNVs/indels between DNase(+) 1203-05 and all DNase(-) ST-19 strains.

1203-05 NODE	NODE position (bp)	Change in DNase(-) ST-19 strains	Product	NEM316 old locus tag [a]
2	5182	G250A (Glu84Lys)	Hypothetical protein	gbs0488
2	14321	T663G (Asn221Lys)	Hypothetical protein	gbs0496
3	42682	264-265insA (Leu89fs) [b]	Nucleoid-associated protein	gbs1791 [b]
4	37354	T926C (Leu309Ser)	Streptococcal histidine triad protein precursor (C-terminal part)	gbs1924 (Y)
4	57207	A591G (Ala197Ala)	NAD-dependent malic enzyme	gbs1906
5	75372	A923T (His308Leu)	Zinc protease	gbs2113
6	85300	T-->C		IGR (gbs1083-gbs1084)
6	102417	C912T (Pro304Pro)	Phosphoglucomutase	gbs1100
8	62062	T-->C		IGR (gbs0923-gbs0924)
9	52087	T317C (Leu106Pro)	YlbF/YmcA family competence/transcriptional regulator	gbs1446
10	5774	T434C (Phe145Ser)	FMN-dependent NADH-azoreductase 1	gbs0259
10	47782	G1511A (Arg504His)	Ribonuclease Y	gbs0295
10	78357	A33C (Ile11Ile)	Ribosome hibernation promotion factor	gbs0326
15	42542	T305C (Ile102Thr)	GTPase Obg	gbs1537
15	49967	C1555G (Leu519Val)	UvrABC system protein B	gbs1531
16	16085	A220G (Ser74Gly)	putative FMN/FAD exporter YeeO	gbs0056
16	30300	T795C (Pro265Pro)	N5-carboxyaminoimidazole ribonucleotide synthase	gbs0043
19	147	T-->C		IGR (gbs1395-gbs1396)
19	13854	A297G (Ter99Trpext*)	TetR family transcriptional regulator	gbs1381
20	38268	2892-2909dupAGCACGTAAAGCTGAAGA (Glu970-Gly971insAlaArgLysAlaGluGlu) [c]	Hypothetical protein	---
20	50146	A630C (Lys210Asn)	Protein DedA	gbs0662
22	7144	A324G (Ala108Ala)	Cystathionine beta-lyase	gbs1283
30	**10716**	**A781G (Lys261Glu)**	**Endonuclease**	**gbs0609 [d]**
45	549	T-->C		IGR (gbs0874-gbs0875)

[a] ORFs designations according to NEM316 genome (GenBank accession number NC_004368.1)
[b] The 1bp insertion leads to a premature stop, yielding a smaller ORF in all ST19 DNase(-) strains and NEM316
[c] The 18bp in-frame insertion yields a longer ORFs in all ST19 DNase(-) strains
[d] Gene putatively encoding secreted DNase [163]
Y Pseudogene

3.3.5. Pan-genomic analyses

We also aimed to identify accessory genes with the potential to discriminate the two strains exhibiting DNase activity exceptions with the respective STs. As such, we performed a pairwise comparison of gene content between DNase(-) ST17 2211-04 and the genetic closest DNase(+) ST17 strain, as well as between the DNase(+) ST19 1203-05 and the genetic closest DNase(-) ST19 strain.

Overall, a total of 2759 genes were detected after pan-genome analysis within all ST17 strains with assembled genome (n=12), with 1788 genes present in more than 99% of the strains. Only 351 genes were classified as cloud genes (present in less than 15% of strains). A zoom-in analysis focusing the ST17 2211-04 and the closest related 1232-06 strain revealed that the proportion of the shared genome was very high (95%, n=2016/2128 genes), with only 112 genes (5%) differentially present through the 2.1 Mb genome of the two strains. However, 64 of these genes were also present in some of the ST17 DNase(+) strains (and in some of the ST19 strains) leaving us with 48 genes found to be exclusive of the accessory content of DNase(-) ST17 2211-04 (Supplementary Table S3.2). Interestingly, these encompass phage-related genes, including genes coding for several hypothetical proteins and for the LexA and the Single-stranded DNA-binding proteins, which are associated with stress response and DNA repair [201-204]. In fact, DNase(-) ST17 2211-04 was found to harbor two intact phages of 35Kb (36.7%GC) and 24.6Kb (40.1%GC) integrated in its genome (Table 3.3). Although the latter is exclusive, not having homology in *S. agalactiae*, it is 96% homolog (99% identity) to *Streptococcus* phage Javan11 (GenBank accession number MK448669.1)

A 34Kb transposon 94% homologous to TnGBS2.3 [205] was also found in DNase(-) ST17 2211-04 strain, carrying several VirB/D4 type IV secretion system (T4SS) and LPXTG protein genes (Supplementary Figure S3.3). It is inserted in-frame within the intergenic region (IGR) between the *gbs1087* and *gbs1088* homologs (which code for the fibronectin-binding protein FbsA and the major cardiolipin synthase ClsA, respectively), being flanked by a perfect 13bp direct repetitive sequence (TTACTTTTTAAAT). However, Blastn analysis revealed that this transposon is also present in DNase(+) ST17 776-06 strain. In contrast to 2211-04, the transposon insertion in 776-06 interrupts the transcriptional terminator region of a glycosyltransferase (*gbs0452* homolog), group 2 family protein, involved in the biosynthesis of *S. agalactiae* Group B carbohydrate [206]. The functional consequences of this insertion are unknown.

Conversely, all ST17 DNase(+) strains were found to uniquely harbor 9 accessory genes (Table 3.3), the majority of which code for hypothetical proteins (n=4) or are involved in cellular processes and signaling (n=2). Two of these exclusive genes encode the Eco47II family restriction endonuclease and the C-5 cytosine-specific DNA methylase, which are components of prokaryotic DNA restriction-modification mechanisms that protect the organism against invasion by foreign DNA [207, 208]. Considering that this accessory gene set is absent in the ST17 DNase(-) 2211-04 as well as in all ST19 strains, it shows up as an important candidate to be involved in the DNases' activity phenotype, besides the recognized DNase coding genes.

Table 3.3 List of accessory genes differentially present between and within ST17 and ST19 strains.

	ST17		ST19	
Feature	**DNase(-) 2211-04**	**DNase(+) (n=14)**	**DNase(+) 1203-05**	**DNase(-) (n=17)**
Phage (24.6Kb) [a]	√			
Transposon (TnGBS2.3 homolog)	√	776-06 [b]		
Cell division protein FtsK		√		
Hypothetical protein 1		√		
Hypothetical protein 2		√		
Hypothetical protein 3		√		
Hypothetical protein 4		√		
Serine/threonine-protein kinase PknD		√		
Eco47II family restriction endonuclease		√		
C-5 cytosine-specific DNA methylase		√		
Tyrosine recombinase XerD		√		
Laminin-binding protein (Lmb)		√		√
Serine-rich repeat protein Srr-1			2 smaller ORFs	√
C5a peptidase		√		√

[a] Phage (24.6Kb) has no homology in *S. agalactiae* but is 96% homolog (99% identity) to *Streptococcus* phage Javan11 (GenBank accession number MK448669.1)
[b] 96% homology (99.96% identity) with 2211-04

For ST19, the pan-genome analysis detected a total of 3073 genes among all ST19 strains with assembled genome (n=12), of which 1713 were shared by more than 99% of the strains and 878 genes were classified as cloud genes (present in less than 15% of strains). A fined-tune analysis focused on the weak DNase(+) 1203-05 and one of the closest DNase(-) ST19 strains (461-05, 724-06 or 1434-05) revealed that each pair shared at least 93% of the genes (with only 85-130 genes being differentially present). While no accessory content was found to be exclusively present in 1203-05, six genes are uniquely harbored by all ST19 DNase(-) strains (data not shown). However, they are also present in all ST17 strains (including in DNase(-) 2211-04), suggesting that they do not contribute to the DNases' activity phenotype.

3.3.6. Virulome characterization

In order to identify putative virulence determinants in *S. agalactiae* that may explain the observed DNase activity exceptions, genomes were queried against a custom database of genes potentially linked to adaptation and pathogenicity. While no specific virulence traits were identified for ST17 DNase(-) 2211-04 when compared with the other ST17 strains, some exclusive particularities were found for ST19 DNase(+) 1203-05 within the ST19 group (Supplementary Table S3.3), but it is unlikely that they may contribute for the observed DNase activity phenotype. For instance, both laminin-binding protein (*lmb*) and C5a peptidase (*scpB*)

genes, which disrupt complement-mediated innate immunity [24], were found to be absent in the weak DNase(+) strain. However, this is intriguing because the surface C5a peptidase is considered to prevent the recruitment of neutrophils to infection sites and, consequently, preclude the production of NETs [24, 108, 175]. Additionally, two apparent functional smaller serine-rich repeat protein *srr-1* ORFs were also identified in 1203-05, instead of an intact gene (whose encoded protein is linked to host extracellular matrix) characteristically exhibited by all ST19 strains.

CRISPR-Cas analysis (Supplementary Figure S3.4) was not informative in elucidating the disparities of the observed DNase activity phenotypes as it correlates with the core-genome SNV-based phylogeny (Supplementary Figure S3.1), with both the DNase(-) 2211-04 and the closest DNase(+) 1232-06 grouping together (cluster 3) in ST17; regarding the weak DNase(+) 1203-05, it confirmed its higher closeness to the related DNase(-) ST19 1434-05, 461-05 and 724-06 strains.

3.4. Discussion

The present work aimed to get insight into the genetic background of DNase production in *S. agalactiae* strains, through WGS-based analyses of clinical strains from carrier and infection, belonging to ST17 and ST19. In 2018, Florindo and co-authors [128] found strains displaying a DNase(-) phenotype belonging to a single genetic lineage, CC19, and all CC17 strains displaying a DNase(+) phenotype. As ST17/capsular-type III strains have been associated with LOD and neonatal meningitis [32, 68, 123, 124], these findings seemed to corroborate the hypothesis of a major role of this enzyme in hypervirulence, namely in the evasion processes from the host defense mechanisms [128]. In the present study, all ST17 strains, except one (DNase(-) ST17 2211-04), were found to be DNase(+), while strains belonging to ST19 were classified as DNase(-), with the exception of strain DNase(+) ST19 1203-05, which revealed some DNase activity. These exceptions in DNase activity from both STs thus constituted an important opportunity to track the genetic basis underlying the DNase production.

While the virulome characterization was not helpful since it did not allow to present genetic characteristics that distinguished each of the exception strains from the ST group they belong to, it was interesting to verify that the analysis of the genetic variability of the seven genes putatively encoding secreted DNases [163] provided some clues. In fact, we found an exclusive non-synonymous alteration (Glu261Lys) in *gbs0609* for strain DNase(+) ST19 1203-05, although it was not possible to determine the structural effect of this amino acid substitution. On the other hand, a non-synonymous alteration (Thr10Ala) in *gbs0661*, coding for the major *S. agalactiae* nuclease (NucA) [163], was found in strain DNase(-) ST17 2211-04. Curiously, it involves the

substitution of a polar uncharged residue by an amino acid with a hydrophobic side chain within the hydrophobic domain of the predicted signal peptide, which is required for export/secretion of the mature protein [163, 209]. Although an hydrophobicity increase was showed to favor protein processing and translocation [209, 210], we observed an opposite phenotypic scenario, so we speculate whether the observed mutation affecting NucA signal peptide may have impaired the protein secretion, leading to DNase(-) phenotype. In another perspective, we did not found any difference in the predicted NucA mature protein among all ST17 and all ST19 strains, regardless of the DNase activity phenotype. When compared to the well-studied NEM316 NucA, they only differ at one position (Pro99Leu) (as the additional nucleotide variation seen is synonymous), containing a conserved H-N-N motif (H^{148}, $N^{167\ or\ 170}$ and N^{179}) and other key amino acid residues (like Glu^{193}) that are crucial for enzyme activity and metal binding [200]. Therefore, the allelic difference between all these proteins and the NEM316 NucA mature protein seems not justify any differences in the DNase activity phenotype, since some ST17 strains presented a strong DNase activity, similar to NEM316. Accordingly, it is very likely that, besides NucA and other predicted DNases, other proteins, namely those involved in their export/secretion, may account for the DNase phenotypic differences observed between ST17 and ST19 strains.

The analysis of the core-genome identified some specificities (SNVs or indels) differentiating the DNase(-) ST17 2211-04 and the DNase(+) ST19 1203-05, from all DNase(+) ST17 strains and all DNase(-) ST19 strains, respectively. We thus looked for a rationale that could justify why some of the altered genes could potentially be implicated in DNase release or activity. For instance, ATP-dependent Clp protease ATP-binding subunit ClpC and ClpX, the molecular chaperone DnaK and the class I SAM-dependent methyltransferase were mutated in the DNase(-) ST17 2211-04 and are involved in metabolism, stress response and pathogenesis [211-218]. Genes related to bacterial stress response and DNA repair were mutated in DNase(+) ST19 1203-05, namely genes encoding for the GTPase Obg and the excinuclease ABC subunit B. GTPase Obg has been associated with a variety of cellular functions, including regulation of the cell stress response and persistence in response to nutrient starvation, while the excinuclease ABC subunit B is pivotal for stress responses, and recognition and processing of DNA lesions [219, 220]. For DNase(+) ST19 1203-05 strain, the gain of a premature stop codon in the TetR/AcrR family transcriptional regulator gene could play an important role in metabolic regulation. This family of transcriptional regulators is widely associated with antibiotic resistance and with the regulation of genes encoding small-molecule exporters [221-223]. Moreover, the ST19 DNase(+) strain presented a core-SNV within a non-coding sequence of a threonine/serine exporter family protein, which hypothetically may influence the expression profile. Interestingly, serine/threonine signaling cascades were linked with virulence, namely in *S. agalactiae* [224]. However, the impact of these mutations in DNase enzymatic performance remains undetermined. In fact,

enzymes act under of multiple levels of regulation and, in order to clarify the overall picture of these regulation processes, different levels of "omics" data such as transcriptomics, proteomics, metabolomics and epigenomics should be integrated [225, 226].

Concerning the pan-genome, the analysis was focused on genes exclusive of the 'exception strains' for DNase activity (DNase(-) ST17 2211-04 and DNase(+) ST19 1203-05). A total of 48 phage-related genes were found to be exclusive of DNase(-) ST17 2211-04. Although it was not possible to determine the exact integration coordinates in the genome, this strain was found to harbor a unique intact phage with no homology in available *S. agalactiae* genome sequences. In a pure speculative basis, this could somehow contribute to the DNase phenotype of this strain. In fact, phages were found to regulate bacterial populations by altering bacterial gene expression, as well as to contribute for the evolution of bacterial hosts (through gene transfer) and to bacterial pathogenesis [227, 228]. The pan-genomic analysis also allowed the identification of a transposon homologous to TnGBS2.3 in two ST17 strains (DNase(-) 2211-04 and DNase(+) 776-06), yet with different sites of insertion. TnGBS transposons are common among *S. agalactiae* isolates and carry several VirB/D4 T4SS and LPXTG protein genes [205]. While VirB/D4 T4SS mediates the conjugative transfer of DNA, enhances bacterial pathogenicity and was proposed to mediate largescale infectious outbreaks in humans [205, 229], LXPTG have been shown to be involved in host colonization, biofilm formation and immune modulation [230, 231]. The site of insertion of the transposon in DNase(-) ST17 2211-04 strain is proximal to genes belonging to a specialized system for excreting extracellular proteins across bacterial cell membranes, the type VII protein secretion system, which has been associated with virulence in *Staphylococcus aureus* [232]. Although the transference of new genes to *S. agalactiae* genome is thought to modify the expression of neighboring genes at the integration site [233], appropriate transcriptomic analysis is required to evaluate the impact of this transposon insertion.

The identification of nine accessory genes among all ST17 DNase(+) strains could also be related with the phenotypes observed for DNase production. Among these genes there are two encoding for proteins involved in mechanisms that protect the organism against invasion by foreign DNA [207], the Eco47II family restriction endonuclease and the C-5 cytosine-specific DNA methylase. The observation that none of these loci was found in any DNase(-) strain may suggest a contribution of these proteins to the lack of DNase activity. Nevertheless, this hypothesis also warrants further investigation through functional assays.

In conclusion, we globally characterized the genome of *S. agalactiae* ST17 and ST19 strains, isolated from invasive and carriage clinical presentations, DNase(+) and DNase(-), in an attempt to shed some light on the genetic background underlying *S. agalactiae* DNase activity. Some loci and SNVs / indels potentially contributing to DNase activity were identified. However, these data require further research to test their robustness, starting by evaluating *S. agalactiae*

strains of different ST and diverse DNase production phenotypes, in order of reaching a better understanding of the pathways whereby DNases contribute to *S. agalactiae* pathogenesis.

Acknowledgments

The authors wish to thank Barbara Spellerberg for the kind gift of *S. agalactiae* strains from collection of the Institute of Medical Microbiology and Hygiene (IMMH), Ulm University, Germany, and to Ana Tenreiro of the Faculdade de Ciências of University of Lisbon, for the use of the fluorimeter.

This work was supported by the GenomePT project (POCI-01-0145-FEDER-022184) from Fundação para a Ciência e a Tecnologia (FCT).

Chapter IV

Global gene expression analysis of *Streptococcus agalactiae* at exponential growth phase

This chapter corresponds to a manuscript with minor changes with the following reference:

Silvestre, I., V. Borges, S. Duarte, A. Nunes, R. Sobral, L. Vieira, J. P. Gomes, M. J. Borrego. 2020. DOI: 10.1101/2020.11.13.381939

Personal contribution

IS performed the experiments, co-interpreted the data, and wrote the paper.

Abstract

Streptococcus agalactiae is a leading cause of neonatal infections and an increasing cause of infections in adults with underlying diseases. One of the first *S. agalactiae* isolates to be subjected to whole genome sequencing was NEM316, a strain responsible for a fatal case of septicemia that has been widely used as reference strain for *in vitro* assays. Whole transcriptome analyses may provide an essential contribute to the understanding of the molecular mechanisms responsible for bacteria adaptation and pathogenicity, still, so far, very few studies were dedicated to the analysis of global gene expression of *S. agalactiae*. Here, we applied RNA-sequencing to perform a comparative overview of the global gene expression levels of the *S. agalactiae* reference strain NEM316 at the exponential growth phase. Genes were ranked by expression level and grouped by functional category and 46% of the top-100 expressed genes encode proteins involved in "Translation, ribosomal structure and biogenesis". Among the group of highly expressed genes were also represented genes with no assigned functional category. Although this result warrants further investigation, most of them might be implicated in stress response. As very little is known about the molecular mechanisms behind the release of DNase's *in vitro* and *in vivo*, we also performed preliminary assays to understand whether direct DNA exposure affects the gene expression of strain NEM316 at the exponential growth phase. No differentially expressed genes were detected, which indicates that follow-up studies are needed to disclose the complex molecular pathways (and stimuli) triggering the release of DNase's. In general, we provide data on the global expression levels of NEM316 at exponential growth phase that may contribute to better understand *S. agalactiae* adaptation and virulence.

4.1. Introduction

Streptococcus agalactiae belongs to the family of streptococcaceae and is a common inhabitant of the healthy human gut and urogenital tract, particularly in women [147]. As an opportunistic pathogen, *S. agalactiae* is a major cause of neonatal infections and mortality, and has also emerged as a pathogen in elderly and immunocompromised adults [139]. *S. agalactiae* is capable to adhere to various host cell types, namely epithelial cells of the vagina and the lung, endothelial cells and micro-vascular endothelial cells of the blood-brain barrier [234]. However, the molecular mechanisms underlying the transition from colonization to infection have yet to be disclosed, and while the exact pathogenic features of *S. agalactiae* remain unrevealed, many virulence factors have been proposed to explain infection-related pathogenesis [128]. In *S. agalactiae* reference strain NEM316 (genotype: III/ ST23) [11], isolated from an infected infant, several extracytoplasmic virulence factors have been identified, such as capsule, proteases, adhesins, haemolysin, pili and pigment. These factors mediate adhesion and epithelial cell invasion, and/or antagonize the immune system during phagocytosis [129, 147].

Understanding the transcriptomic setting allows to obtain insights on the pathways of bacterial physiology, metabolism, and adaptation to changing environments [117, 235]. Over the past decade, RNA-seq has become an indispensable tool for transcriptome-wide analysis of differential gene expression. Together with improved computational tools for data analysis, innovations in RNA-seq technologies are contributing to a better understanding of RNA biology and intermolecular interactions that govern RNA function, as well as transcriptional dynamics changes driving bacterial response and adaptation, to distinct growth conditions and external stimuli [236]. Recently, RNA-seq technology has been successfully applied to clarify some molecular mechanisms of pathogenesis in *S. agalactiae* [110, 116-121]. The first *S. agalactiae* comparative transcriptomic study evidenced several genetic factors (in particular related to the lactose metabolism) likely important in adaptation to the bovine environment [116]. In 2018, Hooven *et al.* [120] identified the gene products necessary for *S. agalactiae* survival in human whole blood, and Cook *et al.* [121] identified novel genes involved in murine vaginal colonization by *S. agalactiae*. The contribute of CRISPR-associated protein-9 to *S. agalactiae* colonization and disease was also investigated by RNA-seq [119]. While these targeted studies provided insightful data about specific *S. agalactiae* adaptive traits, the expression levels of reference strains during their normal growth in the laboratory have not yet been systematized. In this study, we investigated the global gene expression of *S. agalactiae* NEM316 during the exponential growth phase by RNA-seq. We also performed preliminary assays to understand whether direct DNA exposure affects *S. agalactiae* gene expression.

4.2. Materials and methods

4.2.1. Whole genome sequencing of the laboratory reference strain NEM316

All experiments were conducted using the *S. agalactiae* reference strain NEM316 (ATCC 12403; genotype: III/ ST23). The reference strain NEM316, maintained in laboratory at -80ºC in cryopreservation tubes (Cryoinstant Red, VWR, Belgium), was cultured and submitted to WGS as previously described in chapter III. The reads were deposited in the ENA (Bioproject PRJEB41294) under the accession number ERR4836035.

In order to evaluate the genetic differences between the genome of our laboratory passaged reference strain and the publicly available genome (GenBank accession number NC004368), two strategies were applied: i) de novo genome assembly using INNUca v.4.0.1 (https://github.com/B-UMMI/INNUca) [182], followed by genome alignment and inspection using MAUVE (http://darlinglab.org/mauve/mauve.html) (to inspect for the presence of structural changes, such as large indels); ii) reference-based mapping using Snippy v3.2 (to detect SNPs and small indels) (https://github.com/tseemann/snippy).

Clusters of Orthologous Groups (COGs) categories were assigned to the amino acid sequences retrieved from the NEM316 NCBI annotation (GenBank accession number NC004368) using "cdd2cog" script [237] after RPS-BLAST+ (Reverse Position-Specific BLAST) (e-value cut-off of 1e-2), where only the best hit (lowest e-value) and first COG were considered.

4.2.2. Bacterial culture for RNA-seq

Bacterial clones of NEM316 were grown in Columbia agar supplemented with 5% sheep blood (Biomérieux, Marcy l'Etoile, France) with 5% CO_2 for 24 h and then inoculated in Todd Hewitt selective media broth (THB) (BD®, New Jersey, USA) that were allowed to incubate without shaking at 37ºC, 5% CO_2. Cell growth of *S. agalactiae* strain NEM316 was monitored by OD at 600nm (OD_{600}). At OD_{600}=0,6 (exponential growth phase, see Supplementary material Figure S4.1) 1 ml of bacterial cells were collected by centrifugation (3000 rpm, 10 min), resuspended in 1,8 ml of fresh THB and incubated for 0, 10 and 20 min at 37ºC in the presence of 200µl of PBS (used as control). For direct DNA exposure assays, the same procedure was performed, with the exception that 2µg/ml of DNA (human DNA extracted from Hela cells, used

as a stimulus) was added instead of PBS, and nuclease reaction was stopped by adding EDTA (0.5 M, pH 8.0) at 4ºC. For both conditions, 1 ml of each bacterial culture was collected and immediately subjected to high-speed centrifugation (14000 rpm) for 10 min at 4ºC for RNA extraction. Note that we intentionally did not treat the bacterial culture from which RNA would be extracted with RNAprotect™ Bacteria Reagent (Qiagen, California, USA) because a preliminary assay showed that this product degrades *S. agalactiae* RNA (data not shown).

4.2.3. RNA extraction

RNA was extracted as previously referred [180], with minor changes. Briefly, the cells were lysed in 200 µl of Tris-EDTA buffer, pH 8.0, containing 10 U mutanolysin (Sigma-Aldrich, St. Louis, USA) and 15 mg/ml lysozyme (Sigma-Aldrich, St. Louis, USA), at 37ºC during 90 min. The RNeasy mini kit (Qiagen, California, USA) was used according to manufacturer's instructions. Residual contaminant DNA was removed using 30 U RNase-free DNase (Qiagen, California, USA), and elution was performed with 40 µl of RNase-free water. RNA yield and purity were determined by absorbance measurement at 260 and 280 nm using the Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA). Extracted RNA was finally stored at -80ºC until use.

4.2.4. Bacterial mRNA preparation/purification

Bacterial mRNA was enriched using the Ribo-Zero™ rRNA Removal Kit (Illumina, California, USA) which removed abundant 16S and 23S rRNA from total RNA. The obtained bacterial mRNA was concentrated to a final volume of 14 µl using the RNeasy® MinElute® Cleanup Kit (Qiagen, California, USA). The yield and integrity of the enriched mRNA samples was assessed with an Agilent Bioanalyzer, where the absence of rRNA readings is indicative of the success of rRNA depletion and purity of mRNA.

4.2.5. RNA-seq

Bacterial mRNA-enriched samples were subjected to library construction by TruSeq Stranded mRNA sample preparation kit (Illumina, California, USA). The obtained cDNA

libraries were subjected to RNA-seq on a high-throughput MiSeq Illumina apparatus, targeting around 4 M reads per 1 Mbp. Sequence reads (2x75bp) were subjected to quality control and subsequently mapped to the *S. agalactiae* NEM316 genome (as obtained above) using Bowtie2 [238]. Relative gene expression was quantified and normalized as fragments per kb of CDS per million mapped reads (FPKM) using the Cufflinks software (version 2.1.1; http://cufflinks.cbcb.umd.edu/). For comparative global gene expression analyses between normal growth condition and the "direct DNA exposure", we applied HTSeq-count (https://htseq.readthedocs.io/en/release_0.11.1/count.html#) for read counting and state-of-the-art software for differential expression analysis (namely, EdgeR and Voom/Limma) using the interactive web-tool DEGUST (https://degust.erc.monash.edu/) [239].

The reads were deposited in ENA (Bioproject PRJEB41294) under the accession numbers ERR4836029, ERR4836030, ERR4836031, ERR4836032, ERR4836033, ERR4836034.

4.3. Results and discussion

4.3.1. NEM316 whole genome sequencing

S. agalactiae reference strain NEM316 was isolated 18 years ago [11] and, since then, has been maintained in laboratory. As such, in order to prepare RNA-seq assays, NEM316 was subjected to WGS to evaluate whether this laboratory passaged strain had significant genetic changes in comparison with the publicly available genome [11] (GenBank accession number NC004368). Six genome-dispersed mutations were detected, including three small indels and three single nucleotide polymorphisms (SNPs), corresponding to three non-synonymous mutations (Table 4.1). Among these, we highlight a SNP in *relA*, which encodes an enzyme known to be involved in stringent response and bacterial adaptation to environmental stress [120]. In *S. agalactiae*, *relA* knockout strains demonstrated decreased expression of β-hemolysin/cytolysin, an important cytotoxin implicated in facilitating invasion [120]. Although the impact of these particular mutations at the transcriptomic level is unknown, we cannot rule out the possibility that they reflect events of laboratory adaptation. Notwithstanding, we consider a good practice to analyze the genome backbone of strains subjected to gene expression (or other *in vitro*) assays, as a means to provide more complete data required to better interpret and discuss the results.

Table 4.1 Genomic alterations for *S. agalactiae* NEM316 maintained in laboratory.

Mutation location [a]	Type	Locus tag [b]	Old locus tag [b]	nt (aa) change [c]	Effect	Product
1313729	insertion	GBS_RS06735. pseudogene	gbs1273	811insA[d]	--- [d]	glucose-1-phosphate thymidylyltransferase
1363940	insertion			CTT > CTTT		
1741404	deletion			GATATATA > GATATA		
1839016	SNP	GBS_RS09285	gbs1779	A1077T (Glu359Asp)	Missense variant	Major facilitator superfamily transporter
2001001	SNP	GBS_RS10040	gbs1928	G1952C (Trp651Ser)	Missense variant	bifunctional (p)ppGpp synthetase/guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase (relA)
2129967	SNP	GBS_RS10675	gbs2055	A11T (Lys4Met)	Missense variant	arginine repressor

[a] Polymorphism location refers to the location in the reference genome: NEM316 (GenBank accession number NC004368)
[b] ORFs designations according to reference genome: NEM316 (GenBank accession number NC004368)
[c] The nucleotide changes in open reading frames are presented in the 5' to 3' direction
[d] This locus is a pseudogene in the reference genome (GenBank accession number NC004368) due to a one bp deletion as such NEM316 of our laboratory detains the original not truncated allele

4.3.2. NEM316 transcriptomic analyses

The main goal of the present study was to evaluate the global gene expression dynamics of NEM316 *S. agalactiae* strain during the exponential phase using the RNA-seq technology. This growth phase is particularly interesting when studying the transcriptional activity because most cells in the population are actively dividing and this ensures that the expression of most of the *S. agalactiae* genes is assessed. As such, three time points (0, 10 and 20 minutes) were evaluated during the exponential growth phase in THB.

Firstly, we ranked the *S. agalactiae* genes by expression level and correlated them with the gene functional category (Figure 4.1) (Supplementary material Table S4.1). Huge differences in the median expression levels were observed between different gene functional categories (Figure 4.1), with the top expressed functional category ("Translation, ribosomal structure and biogenesis") revealing a median expression value that was 55-fold higher than the less expressed functional category ("Cell motility"). The top three most expressed functional categories were "Translation, ribosomal structure and biogenesis", "Energy production and conversion" and "Posttranslational modification, protein turnover, chaperones". This result may not be surprising as during the exponential growth phase, the cell division rate is maximum and this implicates a high demand for proteins playing a role in translation and metabolism. The three functional categories with the lowest median expression levels were "Cell motility", "Intracellular trafficking, secretion, and vesicular transport" and "Not assigned/Function unknown".

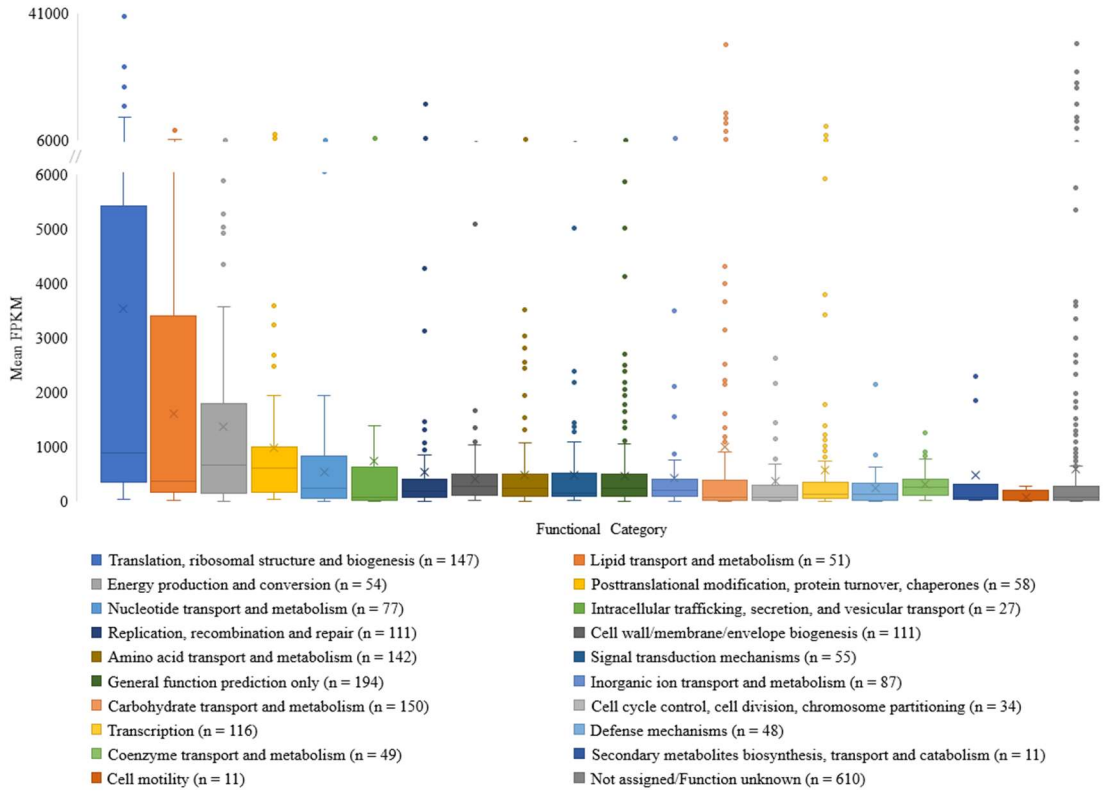


Figure 4.1 Global gene expression by functional categories. Box plot showing the distribution of gene expression levels by functional category, for reference strain NEM16 at the exponential growth phase. Values reflect the mean expression level evaluated at three time points (0, 10 and 20 min) during exponential growth. Genes with mean expression levels < 1 FPKM were excluded from the analysis.

The analysis of the top-100 genes with highest level of expression at the exponential growth phase (Table 4.2) showed that the genes belonging to the functional category "Translation, ribosomal structure and biogenesis" were highly represented (46%), with the proportion of genes from the remainder functional categories never exceeding 15%. Also, 15 genes with unknown function were detected among the highly expressed genes (GBS_RS11490, GBS_RS06375, GBS_RS06390, GBS_RS06380, GBS_RS06405, GBS_RS06400, GBS_RS06385, GBS_RS06800, GBS_RS11205, GBS_RS03445, GBS_RS10615, GBS_RS06395, GBS_RS00250, GBS_RS05190, GBS_RS11525). Although these genes have not been grouped into any functional category by RPS-BLAST against the COG database, fine-tune evaluation of their putative function based on the new NEM316 genome annotation (released on June 2020), plus literature surveys, provided some clues that might justify the observed high expression level. This is the case of GBS_RS00250, which is believed to be required for *S. agalactiae* cell division due to its potential role in peptidoglycan cleavage, since it includes a CHAP domain that has been associated with peptidoglycan hydrolysis [240]. Disruption of this gene was shown to cause an altered cell morphology and an increased susceptibility towards different antibiotics, namely β-lactam antibiotics [241, 242]. GBS_RS05190, a PASTA domain-containing protein, may also be

involved in bacterial cell division as PASTA repeats are known to be key regulators of the membrane during bacterial cell division [243]. GBS_RS11205 is a putative holin-like toxin, and holins, which are encoded by phages, have been considered responsible for disruption of the cytoplasmic membrane to assist endolysins during cell lysis [244]. Although most of these proteins are implicated in the bacterial stress response, like GBS_RS06375, GBS_RS06380, GBS_RS06405, GBS_RS06400, GBS_RS06385, GBS_RS03445 and GBS_RS06395 [116, 245-247], others (GBS_RS11490, GBS_RS06390, GBS_RS06800, GBS_RS10615, GBS_RS11525) do not have any assigned function, neither any predicted functional domain.

The functional category "Intracellular trafficking, secretion, and vesicular transport", which belongs to one of the functional categories with the lowest median expression levels, was also represented among the top-100 most expressed genes by GBS_RS09920 and GBS_RS00560 (preprotein translocase subunit YajC and preprotein translocase subunit SecY, respectively).

The functional category "Signal transduction mechanisms" was represented among the top-100 most expressed genes by only one gene, GBS_RS08755, a response regulator transcription factor, CovR, known to play a role on virulence gene expression [34, 248].

Table 4.2 Top-100 most expressed genes by functional category.

Locus tag [a]	Old locus tag [a]	Product [b]	Functional Category [c]	Mean FPKM [d]
GBS_RS04310	gbs0782	Elongation factor Tu	Translation, ribosomal structure and biogenesis	40081
GBS_RS11490	---	Not assigned	Not assigned/Function unknown	36837
GBS_RS06375	gbs1202	Asp23/Gls24 family envelope stress response protein	Not assigned/Function unknown	32678
GBS_RS09445	gbs1811	Type I glyceraldehyde-3-phosphate dehydrogenase	Carbohydrate transport and metabolism	32335
GBS_RS09575	gbs1836	50S ribosomal protein L34	Translation, ribosomal structure and biogenesis	26121
GBS_RS06390	gbs1205	DUF2273 domain-containing protein	Not assigned/Function unknown	24824
GBS_RS06380	gbs1203	CsbD family protein	Not assigned/Function unknown	22768
GBS_RS06405	gbs1208	GlsB/YeaQ/YmgE family stress response membrane protein	Not assigned/Function unknown	21825
GBS_RS11120	---	50S_ribosomal_protein_L36	Translation, ribosomal structure and biogenesis	20731
GBS_RS06400	gbs1207	GlsB/YeaQ/YmgE family stress response membrane protein	Not assigned/Function unknown	20358
GBS_RS03205	---	HU family DNA-binding protein	Replication, recombination and repair	15942
GBS_RS06385	gbs1204	Asp23/Gls24 family envelope stress response protein	Not assigned/Function unknown	15927
GBS_RS00885	---	50S ribosomal protein L28	Translation, ribosomal structure and biogenesis	15451
GBS_RS00525	gbs0071	Type Z 30S ribosomal protein S14	Translation, ribosomal structure and biogenesis	15276
GBS_RS03485	gbs0608	Phosphopyruvate hydratase	Carbohydrate transport and metabolism	13436

GBS_RS06800	---	DUF3042 family protein	Not assigned/Function unknown	12964
GBS_RS00875	gbs0125	fructose-bisphosphate aldolase	Carbohydrate transport and metabolism	12483
GBS_RS11205	---	Putative holin-like toxin	Not assigned/Function unknown	12444
GBS_RS01260	---	30S ribosomal protein S15	Translation, ribosomal structure and biogenesis	12269
GBS_RS04595	gbs0839	Phosphocarrier protein HPr	Carbohydrate transport and metabolism	12076
GBS_RS03445	---	CsbD family protein	Not assigned/Function unknown	11868
GBS_RS00545	gbs0075	30S ribosomal protein S5	Translation, ribosomal structure and biogenesis	11614
GBS_RS03385	gbs0587	50S ribosomal protein L19	Translation, ribosomal structure and biogenesis	11587
GBS_RS00495	gbs0065	50S ribosomal protein L16	Translation, ribosomal structure and biogenesis	11233
GBS_RS10615	gbs2043	DUF1292 domain-containing protein	Not assigned/Function unknown	11174
GBS_RS00510	gbs0068	50S ribosomal_protein L14	Translation, ribosomal structure and biogenesis	11111
GBS_RS00570	gbs0080	Translation initiation factor IF-1	Translation, ribosomal structure and biogenesis	10915
GBS_RS09435	gbs1809	Phosphoglycerate kinase	Carbohydrate transport and metabolism	10578
GBS_RS00535	gbs0073	50S ribosomal protein L6	Translation, ribosomal structure and biogenesis	10415
GBS_RS07230	gbs1374	50S ribossomal protein L7/L12	Translation, ribosomal structure and biogenesis	10230
GBS_RS00515	gbs0069	50S ribosomal protein L24	Translation, ribosomal structure and biogenesis	10060
GBS_RS10665	gbs2053	cold-shock protein	Transcription	9839
GBS_RS00505	gbs0067	30S ribosomal protein S17	Translation, ribosomal structure and biogenesis	9523
GBS_RS00540	gbs0074	50S ribosomal protein L18	Translation, ribosomal structure and biogenesis	9357
GBS_RS00580	gbs0083	30S ribosomal protein S11	Translation, ribosomal structure and biogenesis	9271
GBS_RS05090	---	30S ribosomal protein S20	Translation, ribosomal structure and biogenesis	9175
GBS_RS06395	gbs1206	Alkaline shock response membrane anchor protein AmaP	Not assigned/Function unknown	9153
GBS_RS00575	gbs0082	30S ribosomal protein S13	Translation, ribosomal structure and biogenesis	8859
GBS_RS02085	gbs0337	acetyl-CoA carboxylase biotin carboxyl carrier protein	Lipid transport and metabolism	8621
GBS_RS00585	gbs0084	DNA-directed RNA polymerase subunit alpha	Transcription	8452
GBS_RS04600	gbs0840	Phosphoenolpyruvate-protein phosphotransferase	Carbohydrate transport and metabolism	8379
GBS_RS00490	gbs0064	30S ribosomal protein S3	Translation, ribosomal structure and biogenesis	8315
GBS_RS03355	---	Type B 50S ribosomal protein L31	Translation, ribosomal structure and biogenesis	8098
GBS_RS07870	---	30S ribosomal protein S21	Translation, ribosomal structure and biogenesis	8026
GBS_RS09175	gbs1757	30S ribosomal protein S18	Translation, ribosomal structure and biogenesis	7951
GBS_RS00555	gbs0077	50S ribosomal protein L15	Translation, ribosomal structure and biogenesis	7785
GBS_RS09765	gbs1874	Peroxiredoxin	Posttranslational modification, protein turnover, chaperones	7672
GBS_RS00590	---	50S ribosomal protein L17	Translation, ribosomal structure and biogenesis	7638
GBS_RS00475	gbs0061	50S ribosomal protein L2	Translation, ribosomal structure and biogenesis	7591
GBS_RS07235	gbs1375	50S ribosomal protein L10	Translation, ribosomal structure and biogenesis	7576
GBS_RS00520	gbs0070	50S ribosomal protein L5	Translation, ribosomal structure and biogenesis	7457
GBS_RS04180	gbs0756	PspC domain-containing protein	Transcription	7382

GBS_RS04315	gbs0783	Triose-phosphate isomerase	Carbohydrate transport and metabolism	6956
GBS_RS00460	gbs0058	50S ribosomal protein L3	Translation, ribosomal structure and biogenesis	6902
GBS_RS09450	gbs1812	Elongation factor G	Translation, ribosomal structure and biogenesis	6790
GBS_RS04440	gbs0808	Superoxide dismutase	Inorganic ion transport and metabolism	6587
GBS_RS09920	gbs1904	Preprotein translocase subunit YajC	Intracellular trafficking, secretion, and vesicular transport	6572
GBS_RS09180	gbs1758	Single-stranded DNA-binding protein	Replication, recombination and repair	6519
GBS_RS09770	gbs1875	Alkyl hydroperoxide reductase subunit F	Posttranslational modification, protein turnover, chaperones	6456
GBS_RS00560	gbs0078	Preprotein translocase subunit SecY	Intracellular trafficking, secretion, and vesicular transport	6382
GBS_RS02055	gbs0331	Ketoacyl-ACP synthase III	Lipid transport and metabolism	6364
GBS_RS02105	gbs0341	Acetyl-CoA carboxylase carboxyl transferase subunit alpha	Lipid transport and metabolism	6312
GBS_RS00480	gbs0062	30S ribosomal protein S19	Translation, ribosomal structure and biogenesis	6299
GBS_RS04320	gbs0784	Phosphoglycerate mutase	Carbohydrate transport and metabolism	6283
GBS_RS08050	gbs1534	Amino acid ABC transporter ATP-binding protein	Amino acid transport and metabolism	6171
GBS_RS02095	gbs0339	Acetyl-CoA carboxylase biotin carboxylase subunit	Lipid transport and metabolism	6123
GBS_RS00465	gbs0059	50S ribosomal protein L4	Translation, ribosomal structure and biogenesis	6070
GBS_RS04980	gbs0916	Nucleoside-diphosphate kinase	Nucleotide transport and metabolism	6062
GBS_RS01325	gbs0210	30S ribosomal protein S9	Translation, ribosomal structure and biogenesis	5995
GBS_RS07565	gbs1438	50S ribosomal protein L27	Translation, ribosomal structure and biogenesis	5926
GBS_RS02050	gbs0330	MarR family transcriptional regulator	Transcription	5924
GBS_RS04805	gbs0882	F0F1 ATP synthase subunit epsilon	Energy production and conversion	5896
GBS_RS05130	gbs0946	FAD-dependent oxidoreductase	General function prediction only	5876
GBS_RS00250	gbs0016	CHAP domain-containing protein	Not assigned/Function unknown	5757
GBS_RS00455	gbs0057	30S ribosomal protein S10	Translation, ribosomal structure and biogenesis	5733
GBS_RS02100	gbs0340	Acetyl-CoA carboxylase carboxyltransferase subunit beta	Lipid transport and metabolism	5565
GBS_RS09455	---	30S ribosomal protein S7	Translation, ribosomal structure and biogenesis	5501
GBS_RS00485	gbs0063	50S ribosomal protein L22	Translation, ribosomal structure and biogenesis	5480
GBS_RS09760	gbs1873	30S ribosomal protein S2	Translation, ribosomal structure and biogenesis	5366
GBS_RS05190	gbs0958	PASTA domain-containing protein	Not assigned/Function unknown	5360
GBS_RS02005	gbs0326	ribosome-associated translation inhibitor RaiA	Translation, ribosomal structure and biogenesis	5352
GBS_RS04875	gbs0895	Thiamine pyrophosphate-dependent dehydrogenase E1 component subunit alpha	Energy production and conversion	5285
GBS_RS02080	gbs0336	Beta-ketoacyl-ACP synthase II	Lipid transport and metabolism	5201
GBS_RS07760	gbs1477	PI-2a pilus major subunit PilB	Cell wall/membrane/envelope biogenesis	5099
GBS_RS10710	gbs2062	50S ribosomal protein L32	Translation, ribosomal structure and biogenesis	5094
GBS_RS03295	gbs0569	(S)-acetoin forming diacetyl reductase	Lipid transport and metabolism	5079
GBS_RS02045	gbs0329	Enoyl-CoA hydratase	Lipid transport and metabolism	5075

GBS_RS00470	gbs0060	50S ribosomal protein L23	Translation, ribosomal structure and biogenesis	5075
GBS_RS04800	gbs0881	F0F1 ATP synthase subunit beta	Energy production and conversion	5029
GBS_RS08755	gbs1672	Response regulator transcription factor	Signal transduction mechanisms	5021
GBS_RS07510	gbs1427	KH domain-containing protein	General function prediction only	5012
GBS_RS11525	---	Not assigned	Not assigned/Function unknown	4975
GBS_RS07575	gbs1440	50S ribosomal protein L21	Translation, ribosomal structure and biogenesis	4954
GBS_RS05135	gbs0947	L-lactate dehydrogenase	Energy production and conversion	4926
GBS_RS09460	gbs1814	30S ribosomal protein S12	Translation, ribosomal structure and biogenesis	4883
GBS_RS02060	gbs0332	Acyl carrier protein	Lipid transport and metabolism	4739
GBS_RS01320	gbs0209	50S ribosomal protein L13	Translation, ribosomal structure and biogenesis	4733
GBS_RS09380	gbs1798	30S ribosomal protein S14	Translation, ribosomal structure and biogenesis	4605
GBS_RS00530	gbs0072	30S ribosomal protein S8	Translation, ribosomal structure and biogenesis	4486
GBS_RS04880	gbs0896	Alpha-ketoacid dehydrogenase subunit beta	Energy production and conversion	4394

[a] ORFs designations according to NEM316 genome (GenBank accession number NC004368)
[b] Predicted function according to new genome annotation for NEM316, updated at 07-JUN-2020 (GenBank version NC004368.1)
[c] Functional categories classified by RPS-BLAST against the COG database
[d] tRNA and rRNA were excluded as their expression levels are biased due to mRNA enrichment procedure

Finally, we were interested in assessing the expression levels of the genes of interest (n=47), according to the highlighted throughout this thesis (Chapter III, Supplementary Table S3.4) (Figure 4.2). The set includes the seven genes that putatively encode for secreted DNases (GBS_RS03720, GBS_RS04825, GBS_RS01045, GBS_RS03490, GBS_RS02295, GBS_RS03960, GBS_RS05380) and the genetic PI, that consists of five genes which encode for pilus assembly [154], PI-1 (that plays an important role in evasion from host innate immunity) (GBS_RS03580, GBS_RS03575, GBS_RS03570, GBS_RS03585 and GBS_RS03565) and PI-2a (that is specifically involved in adhesion and biofilm formation) (GBS_RS07760, GBS_RS07745, GBS_RS07765, GBS_RS07755 and GBS_RS07750) [37, 155-157]. The major nuclease, NucA (GBS_RS03720; old locus tag: gbs0661), identified in *S. agalactiae* NEM316 by Derré-Bobillot and co-authors [163], presented higher mean expression values (ranked at position 520º out of a total of 2169 genes evaluated) than the other DNase genes (Figure 4.2). Among the PI genes, PI-2a presented higher mean expression values (ranked between positions 84º and 718º out of a total of 2169 genes evaluated) than PI-1 coding genes (ranked between positions 1058º and 1973º out of a total of 2169 genes evaluated) (Figure 4.2).

Interestingly, three virulence associated genes were ranked within the top-100 most expressed genes, including the PI-2a pilus major subunit PilB (GBS_RS07760) (Figure 4.2). The other two genes code for adhesins associated to the glycolytic pathway for energy metabolism; they were GBS_RS09445 (Glyceraldehyde-3-Phosphate Dehydrogenase, GAPDH) and

GBS_RS03485 (enolase), belonging to "Carbohydrate transport and metabolism" functional category [1, 249]. GAPDH is also thought to be involved in macromolecular interactions and bacterial pathogenesis, enhancing *S. agalactiae* virulence.

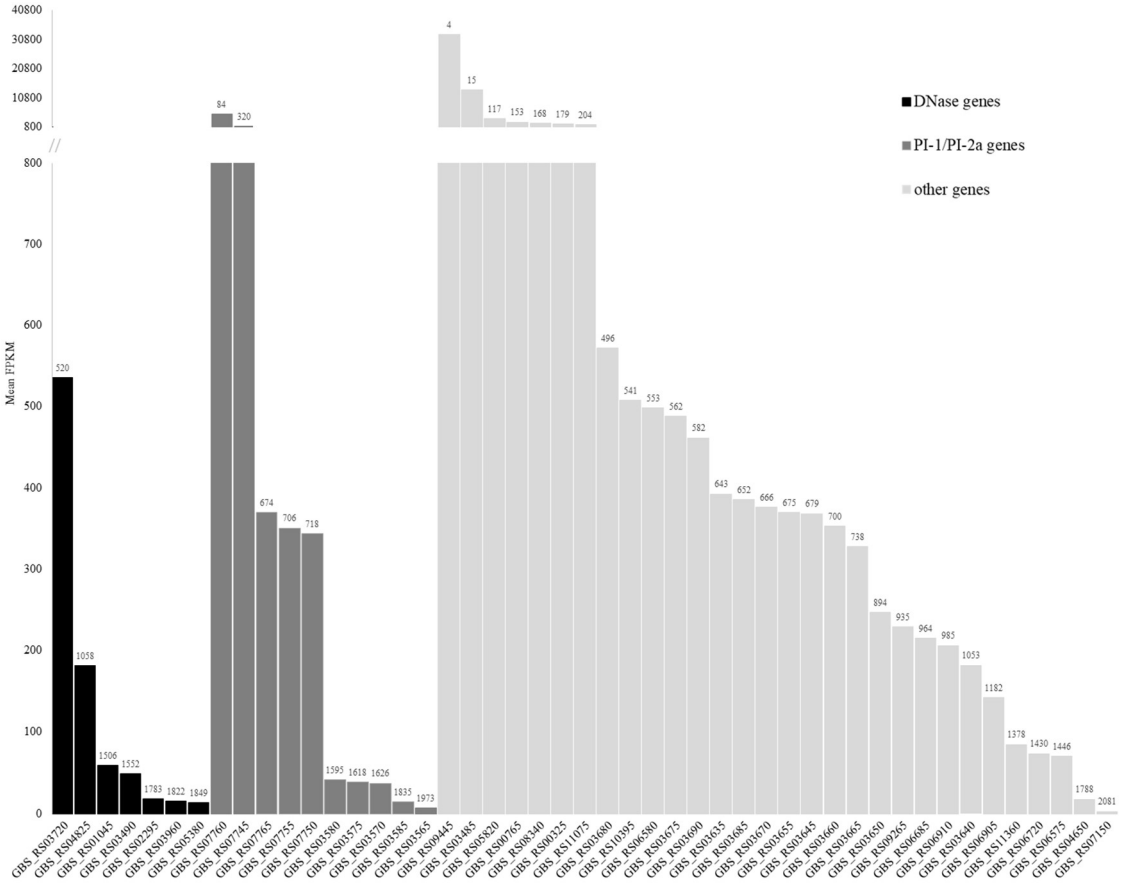


Figure 4.2 Expression level (FPKM units) of genes potentially involved in virulence in reference strain NEM316 at exponential growth phase. Genes are grouped according to their predictive function in three groups, DNase, PI-1/PI-2a and other coding genes, and displayed by increasing level of expression. The values above the bars correspond to the rank position of each gene in the total of 2169 genes evaluated.

4.3.3. NEM316 comparative transcriptomic analysis through exposure to DNA

The scientific interest in *S. agalactiae* extracellular nucleases increased with the discovery, by Brinkmann and colleagues in 2004 [164], that they can disrupt the DNA matrix, which constitutes the nuclear backbone of NETs. Indeed, as NETs (also composed by granule proteins and histones) [163, 177, 250] are released by neutrophils to degrade virulence factors and to kill bacteria, this nuclease-mediated mechanism could play an important role in *S. agalactiae* virulence. Since then, although important knowledge has been acquired about *S. agalactiae* nucleases [128, 138, 163, 200], very little is known about the complex molecular

pathways and stimuli triggering the release of DNase's *in vitro* and *in vivo*. Here, we conducted preliminary RNA-seq assays to understand whether direct DNA exposure could affect gene expression during the exponential growth phase of the DNase producer reference strain NEM316. Although a preliminary assay at the transcriptomic level targeting *gbs0661* revealed no significant expression differences among reference strains NEM316 (DNase producer) and 2603V/R (DNase non-producer) with and without DNA stimuli (data not shown), we hypothesized that exposure to human DNA could trigger differential gene expression in other bacterial genes. These genes could potentially be involved in molecular cascades mediating DNase release and virulence. However, this hypothesis was not verified, as no differentially expressed genes were detected for NEM316 during 20 minutes of the exponential phase (at 37ºC in THB), either with or without the presence of human DNA (interactive online data navigation is available here: http://degust.erc.monash.edu/degust/compare.html?code=b11b5fff2bf525ea465dc450989f351e#) (Figure 4.3). While this data could suggest that human DNA, as a stimulus, has no impact on *S. agalactiae* transcription, there is a need to test other stimuli as well as other strains. In fact, DNases have been shown to be under the control of the extensive regulatory systems in streptococci [138] and further work is required to fully understand the complex regulation of the expression of DNases. Thus, it would be interesting to perform comparative RNA-seq assays involving both high and low DNase producing strains in comparison with DNase non-producers. Also, human neutrophils and NETs could possibly be used as stimuli that better mimic the *in vivo* infection environment. On the other hand, testing *S. agalactiae* strains of human clinical origin, from carriage and invasion, might also be of interest. However, we cannot exclude the putative role of DNases in enzymatic kinetics, i.e., in the catalytic mechanism, and in their activity.

This preliminary assay consolidated our expectation that follow-up studies are needed to disclose the complex molecular pathways (and putative stimuli) triggering the release of DNases. The public release of our data (counts per million, logFC and differential expression statistics, etc) to the scientific community through an interactive and user-friendly web tool (http://degust.erc.monash.edu/degust/compare.html?code=b11b5fff2bf525ea465dc450989f351e#) might further help future analysis and interpretation of other RNA-seq studies in *S. agalactiae*.

Figure 4.3 DEGUST RNA-seq database of NEM316 growing at exponential phase with and without DNA exposure. Interactive web tool showing comparative global gene expression analyses between normal growth condition and the "direct DNA exposure" of NEM316 exponentially growing during 20 minutes.

4.4. Conclusion

This study constitutes the first attempt to systematize the expression levels of *S. agalactiae* reference strain NEM316 during their normal growth in the laboratory, namely during the exponential growth phase. By providing a comprehensive comparison of gene expression by gene functional category and exploring the levels of expression of genes likely associated with virulence, our data may constitute a database for future functional studies. Preliminary data generated for differential gene expression through DNA exposure suggests the need of further studies regarding the impact of NETs in bacteria.

Chapter V

Biofilm formation by ST17 and ST19 strains of *Streptococcus agalactiae*

This chapter corresponds to a manuscript with minor changes with the following reference:

Silvestre, I., M. J. Borrego, and L. Jordão. 2020. Research in Microbiology. 171(8):311-318. DOI: 10.1016/j.resmic.2020.08.001

Personal contribution

IS performed the great majority of the experimental procedures, analyzed and interpreted the data, and wrote the paper.

Abstract

Bacterial biofilms are an important virulence factor with a vital role in evasion from the host immune system, colonization and infection. The aim of the present study was to evaluate *in vitro* the effects of three environmental factors (H^+, glucose and human plasma) in biofilm formation, by carrier and invasive *S. agalactiae* strains of ST17 and ST19 sequence types, including DNase producers and non-producers. Bacteria ability to assemble biofilms was classified based on crystal violet assay. Biofilm formation was also monitored by scanning electron microscopy. Depending on the growth medium used, each bacterial isolate could fit in different biofilm production categories. Our data showed that optimal conditions for *S. agalactiae* biofilm assembly were reached after 48 h incubation at pH 7.6 in the presence of glucose and inactivated human plasma. In the presence of inactivated human plasma, the biofilm biomass of ST19 strains experienced a higher increase than ST17 strains. The composition of the extracellular polymeric matrix of the three strongest biofilm producers (all from ST17) was accessed by enzymatic digestion of mature biofilms and proteins were shown to be the predominant component. The detailed identification of the extracellular protein components should contribute to the development of new therapeutic strategies to fight *S. agalactiae* infections.

Keywords: *Streptococcus agalactiae*; biofilms; extracellular polymeric substance matrix; DNases

5.1. Introduction

Streptococcus agalactiae is an important human pathogen that colonizes the gastrointestinal and genitourinary tracts, and a prevailing cause of neonatal morbidity and mortality worldwide [123, 168].

The infection ability of *S. agalactiae* relies mainly on three mechanisms: (i) the ability to colonize and cross tissue barriers within the host environment, (ii) the capability to evade from host defense mechanisms, and (iii) the expression of virulence factors [123, 128]. *S. agalactiae* is capable to adhere to various host cell types, namely epithelial cells of the vagina and the lung, endothelial cells and micro-vascular endothelial cells of the blood-brain barrier. One of the most critical steps for a successful colonization of the mucosal surface is adhesion to cells and/or surface host proteins. Several specific *S. agalactiae* surface-exposed determinants have been shown to contribute for the adhesion to vaginal and cervical epithelial cells, namely the capsule, pili and surface proteins, which promote adherence to extracellular matrix constituents such as collagen, fibrinogen, fibronectin and laminin [34, 175, 234].

Biofilms are sessile microbial communities in which bacteria are covered by a self-produced matrix of extracellular polymeric substances (EPS) that provides protection from stress factors (e.g. antibiotics and host immune defenses). Thus, biofilm assembly by *S. agalactiae* could enhance its pathogenicity and virulence. Knowledge on the composition of the EPS would enable the control of bacterial biofilm development and its dispersal. Accordingly, depolymerization and degradation of the EPS by several enzymes have been used to investigate its composition [171, 251-254]. It is now known that the biofilm matrix is composed by numerous polysaccharides, proteins and extracellular DNA (eDNA), in proportions that are specific of each bacteria/environmental conditions. The level of eDNA seems to be controlled by extracellular nucleases, which are also involved in other processes such as the development of the three-dimensional biofilm structure, the detachment from a mature biofilm, and the utilization of eDNA as a nutrient source [255]. DNase production by *S. agalactiae* human strains from different CCs, including CC17 was reported, and suggested an important contribution of DNases to hypervirulence [128].

Several studies followed biofilm formation by different *S. agalactiae* strains *in vitro* [126, 156, 168, 169, 171-174]; however, conflicting data have been reported, namely regarding the biofilm-forming capacity of strains from different serotypes, and the correlation between biofilm formation and the glucose concentration or the pH value [126, 156, 169, 172]. Significantly higher biofilm production was described, under acidic pH conditions, for *S. agalactiae* colonization isolates from pregnant women [126, 172], with the strongest biofilm producing isolates belonging

to ST17 [126]. However, lower biofilm formation at low pH values has also been reported [171, 173, 174].

S. agalactiae ST17 lineage has been reported as particularly virulent while ST19 has been mostly associated with carriage [88, 122]. To explain the leading role of ST17 strains in neonatal meningitis, correlation of *S. agalactiae* invasive capacity with biofilm formation has been proposed, and higher frequencies of strong biofilm producers were found among neonatal strains in comparison to colonization strains [126, 172]. However, invasive strains belonging to ST17 and ST19 lineages have also been determined to be weak biofilm formers, while strains isolated from asymptomatic carriers were revealed to be strong biofilm producers [170]. These contradictory results could be due to different experimental conditions.

Considering the dispersed and often conflicting results concerning biofilm production by *S. agalactiae*, the aim of the present study was to evaluate biofilm formation under standardized experimental conditions. Therefore, several environmental factors (different pH values, 4.5, 7.0 and 7.6; and the presence or absence of glucose and plasma) [37, 130] were tested. *S. agalactiae* strains from colonization and infection, DNase producers and non-producers, belonging to ST17 and ST19 lineages, were evaluated for biofilm formation on an abiotic surface (polystyrene). Scanning electron microscopy (SEM) technique was used to evaluate the biofilm phenotype, and three enzymes, DNase, proteinase K and β-N-acetylglucosaminidase, were used to disperse mature biofilms to investigate the composition of the EPS.

5.2. Materials and methods

5.2.1. Strain collection

S. agalactiae reference strains belonging to different genetic lineages were used: NEM316 (genotype: III/ST23; ATCC 12403; GenBank accession number NC004368); COH1 (genotype: III/ST17; GenBank accession number AAJR01000000); 2603V/R (genotype: V/ST110; ATCC BAA-611; GenBank accession number NC004116).

Thirty-eight clinical *S. agalactiae* strains of human origin (n=18 ST17 and n=20 ST19) were also used. This subset of strains was randomly selected from the collections of the Portuguese National Institute of Health (Instituto Nacional de Saúde Doutor Ricardo Jorge, INSA, IP) and the Institute of Medical Microbiology and Hygiene (IMMH), Ulm University, Germany (Table 5.1). These strains have been previously described [42, 87, 128, 256-259]; however, when recently performing whole genome multi locus sequence typing (wgMLST) of these strains (Chapter III), five were reclassified with new or close ST within the same CC (713-05 and 211/#960, new-ST, CC19; 466-06, new-ST, CC17; 5659-05 and 5590-05, ST109, CC17) (Supplementary material

Table S5.1). Considering the ST similarity, for the purpose of the present study, the five strains were included in the prior ST group.

Table 5.1 *S. agalactiae* clinical strains.

Clonal Complex	Capsular serotype/ genotype	Number of human colonizing strains (*n* = 30)	Number of human invasive strains (*n* = 8)	Number of strains with DNase activity (*n* = 18)
CC17	III-2	14	4	17
	II	0	4	0
CC19	III-1	15	0	1
	V	1	0	0

5.2.2. Biofilm assays

5.2.2.1. Biofilm formation

Biofilm evaluation assays were performed as described by Bandeira and co-authors [260], with minor modifications. Briefly, *S. agalactiae* suspensions were prepared in 0.9% sodium chloride from overnight cultures in 5% sheep blood Columbia agar (BioMérieux®, Marcy l'Etoile, France) and then ten-fold diluted in THB (BD®, New Jersey, USA) or RPMI-glucose (commercially available medium with 4.5 g D-glucose) (Thermo Fisher Scientific®, Massachusetts, USA) to a final concentration of 10^7 CFU/ml. Biofilm experiments were performed in 96-well flat-bottom polystyrene cell culture plates (VWR International, Pennsylvania, USA) using THB with or without 1% glucose at different pH values (4.5, 7.0 and 7.6), and RPMI-glucose. Two-hundred microliters of medium-only (THB or RPMI-glucose) were used as blank. Two incubation periods were tested (24 and 48 h) to identify the optimal conditions for biofilm formation *in vitr*o. Microplates were incubated under static conditions at 37°C/5% CO_2 for 24 or 48 h. After incubation, the OD_{600nm} of each culture was measured to ensure that bacterial cells had reached stationary phase at similar density. Then, each well was vigorously washed thrice with sterile distilled water to remove non-adherent bacteria. The attached bacteria were stained at RT for 15 min with 100 µl of 0.1% crystal violet (CV), washed thrice with distilled water to remove excess dye and allowed to dry at RT. The CV was dissolved in 200 µl of 95% ethanol, and OD_{570nm} was accessed using a SpectraMax 340PC (Molecular Devices, California, USA). Three independent experiments using four replicates were performed.

5.2.2.2. Effect of human plasma on biofilm formation

To test the effect of human plasma on biofilm formation (incubation of 48 h, 37ºC/5% CO_2), plates were prepared:

a) As described in section 5.2.2.1 but replacing THB by RPMI-glucose [130] supplemented with or without 11% or 15% of heat inactivated fetal bovine serum (hiFBS) or inactivated human plasma (IHP) with a fibrinogen concentration of 2.59 g/L.
b) Incubating 96-well flat-bottom polystyrene cell culture plates (VWR International, Pennsylvania, USA) at 37ºC/5% CO_2 for 24 h with 200 µl RPMI-glucose supplemented with 15% IHP before proceeding as described in section 5.2.2.1, using RPMI-glucose.

RPMI-glucose medium was used as blank.

5.2.2.3. Interpretation of the results of biofilm formation

Results were interpreted according to Stepanovic and co-authors [261] and required the definition of a cut-off value that separated biofilm producing from non-biofilm-producing strains. Briefly, the average OD_{570nm} (in this section referred as OD) values were calculated for all tested strains and blank. The cut-off value (ODc) was defined as three standard deviations (SD) above the mean OD of the blank. Based upon the previously calculated OD values (ODs), results were interpreted as follows: ODs ≤ ODc = non biofilm producer (NBP); ODc < ODs ≤ 2 x ODc = weak biofilm producer (WBP); 2 x ODc < ODs ≤ 4 x ODc = moderate biofilm producer (MBP); 4 x ODc < ODs = strong biofilm producer (SBP).

5.2.2.4. Effect of enzymatic treatment on biofilms

EPS composition was investigated using DNase, proteinase K and β-N-acetylglucosaminidase that digest, eDNA, proteins and exopolysacharides, respectively.

The assays of the enzymatic treatment of biofilms were performed as described previously [251], with minor modifications. Briefly, 96-well flat-bottom polystyrene cell culture plates (VWR International, Pennsylvania, USA) with *S. agalactiae* suspensions were prepared as described in Section 5.2.2.1 using a medium volume of 200 µL (THB supplemented with 1% glucose or RPMI-glucose supplemented with 15% IHP); the plates were incubated for 48 h. Non-adherent bacteria were removed by vigorous wash thrice with sterile distilled water, before adding DNase I (100 µg/ml or 1000 µg/ml) (Qiagen, Germantown, United States) or 100 µg/ml

proteinase K (Roche, Penzberg, Germany) solutions in PBS or 1.2 units β-N-acetylglucosaminidase (Sigma, St. Louis, MO, USA) in acetate buffer pH 5.0. Plates were further incubated at 37ºC/5% CO_2 for 1h (DNase I and proteinase K) or 2 h (β-N-acetylglucosaminidase) as previously described [251, 262, 263]. The culture supernatant of reference strain NEM316 at stationary phase, rich in DNases (Supplementary material Figure S5.1), was also tested in a 1:1 dilution in PBS and incubation at 37ºC/5% CO_2 for 1 h. This supernatant was used to evaluate putative specificities of streptococcal DNase that could be correlated with biofilm production. The NEM316 culture supernatant was obtained during stationary phase in THB supplemented with 0.5% yeast extract. The culture was centrifuged (10 min. 1811 x g) (Eppendorf 5810R, Hamburg, Germany) and the supernatant was sterilized by filtration through a 0.2 µm diameter pore filter (VWR International, Pennsylvania, USA). Biofilms incubated with PBS or acetate buffer pH 5.0 were used as negative controls for DNase (DNase I/NEM316 culture supernatant) or proteinase K and β-N-acetylglucosaminidase, respectively. After incubation, wells were washed once with sterile distilled water and stained for 15 min with 100 µL 0.1% CV at RT and further processed for biofilm evaluation as described under 5.2.2.1.

5.2.2.5. Scanning Electron Microscopy

Biofilms assembled on polystyrene for 48 h were processed as previously described [251, 260] for SEM in topographic mode with minor modifications. Biofilms were fixed with 4% paraformaldehyde in PBS (Sigma, St Louis, MO, USA) for 15 min at RT, washed with PBS and post-fixed in 1% osmium tetroxide (EMS, Hatfield, PA, USA) in PBS for 90 min on ice protected from light. Samples were washed twice with PBS and twice with water for 10 min each, before dehydration being performed with 25, 50, 70, 80 and 95% ethanol solutions for 15 min each, and twice with absolute ethanol (Merck) for 30 min at RT. Samples were mounted on the sample holder with carbon tape (EMS), sputter-coated with carbon (20 nm) using a Sputter Coater QISOT ES (Quorum Technologies, Laughton, UK) and analyzed under an electron microscope, JSM-7100F (JEOL, Tokyo, Japan).

5.2.3. Statistical analysis

Statistical analysis of biofilm formation by *S. agalactiae* strains was performed with GraphPad (GraphPad Software, San Diego, CA). The significance of differences in the relative amounts of biofilm produced by the different *S. agalactiae* strains under each test condition was assessed by Paired T test, Wilcoxon matched-pairs signed rank test, Mann Whitney test or Kruskal-Wallis test. Significance was assigned to P values < 0.05.

5.3. Results

5.3.1. Biofilm assays – effect of glucose, pH and human plasma

To define the optimal conditions for *in vitro* biofilm formation, the time of incubation in THB (starting pH, 7.6), in the presence or absence of 1% glucose, was tested. The production of biofilm by *S. agalactiae* strains in THB (supplemented or not with 1% glucose) was significantly higher for 48 h than for 24 h ($P < 0.0001$). A similar variation in biofilm biomass was observed for biofilms formed in THB supplemented with 1% glucose *vs* not supplemented ($P < 0.0001$). All strains produced more biofilm in THB supplemented with 1% of glucose (in comparison with non-supplemented THB), and this increase was higher for strains belonging to the hypervirulent ST17 clone ($P < 0.0001$), namely for strains classified as strong biofilm producers (1234-05; 1392-05; 1532-07).

The ability of the 38 *S. agalactiae* strains to develop biofilm on polystyrene surfaces was also evaluated after 48 h (using THB supplemented with 1% glucose) at different pHs (4.5, 7.0 and 7.6). Firstly, to explore the influence of pH on *S. agalactiae* growth, NEM316, COH1 and 2603V/R reference strains were grown in THB supplemented with 1% glucose at pH 4.5, 7.0 and 7.6. Typical bacterial growth curves were observed at pH 7.0 and 7.6 but not at pH 4.5, for which growth was reduced (data not shown). Biofilm biomass in THB supplemented with 1% glucose at pH 4.5, 7.0 and 7.6 significantly varied ($P < 0.0001$), and a general decrease in biofilm formation was observed under acidic and neutral-pH conditions, in comparison with slightly alkaline pH 7.6. At pH 4.5 and 7.0, independently of the ST, all *S. agalactiae* strains were classified as WBP. At pH 7.6, WBP and MBP strains were identified for both STs, but only ST17 included strains classified as SBP (Table 5.2). In fact, 61% of ST17 strains were classified either as MBP or SBP (11/18) and 39% as WBP (7/18), contrasting with only 40% of MBP (8/20) and 60% of WBP (12/20) for ST19 strains. A statistically significant difference ($P = 0.044$) between biofilm formation by ST17 and ST19 strains was only verified at pH 7.6, once again with higher increase of OD_{570nm} for the three SBP strains (1234-05; 1392-05; 1532-07).

Table 5.2 48 h-old biofilms of *S. agalactiae* ST17 and ST19 strains in THB supplemented with 1% glucose at pH 7.6.

Sequence type	Biofilm assembly category [a]	Isolates (n / %)
ST17 (n=18)	Week	7 (38.9 %)
	Moderate	8 (44.4 %)
	Strong	3 (16.7 %)
ST 19 (n=20)	Week	12 (60.0 %)
	Moderate	8 (40.0 %)
	Strong	0

[a] Biofilm assembly was classified according to Stepanovic and colleagues [261]

In order to establish experimental conditions that resemble the vaginal environment and to investigate the role of fibrinogen in biofilm formation by *S. agalactiae*, cell culture medium supplemented with IHP was also tested. Firstly, the incubation time (24 or 48 h) in RPMI-glucose medium was tested, and the amount of biofilm formed for each ST at both incubation periods was not significantly different ($P = 0.501$) (Table 5.3). Biofilm formation by the three reference strains (NEM316, COH1 and 2603V/R) was therefore evaluated using RPMI-glucose supplemented with 11% or 15% of hiFBS or IHP, after a 48 h incubation period. In RPMI-glucose supplemented with hiFBS (11% or 15%) all *S. agalactiae* strains formed a low amount of biofilm and the three reference strains formed denser biofilms in the presence of IHP than in RPMI-glucose medium only or RPMI-glucose supplemented with hiFBS. In fact, biofilm formation increased proportionally to the IHP amount added to RPMI-glucose (Figure 5.1).

Table 5.3 Classification of biofilm assembly by ST17 and ST19 *S. agalactiae* strains after 24 and 48h in RPMI-glucose.

Sequence type	Biofilm assembly category [a]	Isolates (n / %)	
		24 h	48 h
ST17 (n=18)	Week	8 (44.4%)	7 (38.9%)
	Moderate	7 (38.9%)	8 (44.4%)
	Strong	3 (16.7%)	3 (16.7%)
ST 19 (n=20)	Week	4 (20%)	2 (10%)
	Moderate	14 (70%)	16 (80%)
	Strong	2 (10%)	2 (10%)

[a] Biofilm assembly was classified according to Stepanovic and colleagues [261]

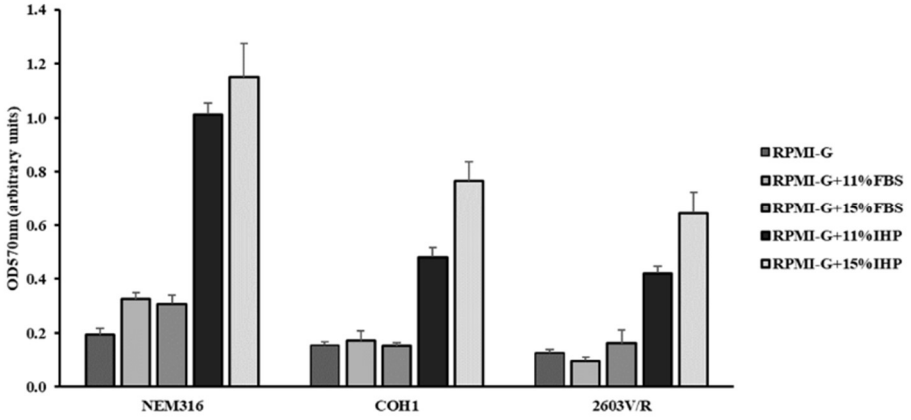


Figure 5.1 Biofilm formation of *S. agalactiae* reference strains NEM316, COH1 and 2603V/R. *S. agalactiae* reference strains were allowed to assemble biofilms in RPMI-glucose medium only or with addition of either 11% or 15% of hiFBS or IHP for 48 h. Biofilm biomass was determined by CV assay.

Based on these results, the RPMI-glucose medium supplemented with 15% IHP was selected to evaluate biofilm formation by ST17 and ST19 *S. agalactiae* strains (at 48 h, to standardize the time of incubation). The effect of IHP on biofilm assembly was tested for two conditions: i) IHP added directly to the medium or ii) added to the medium and used to coat the polystyrene surface for 24 h before the biofilm assay. Coating of the polystyrene plates with RPMI-glucose medium supplemented with 15% IHP during 24 h at 37ºC/5% CO_2 seemed to decrease biofilm formation. Indeed, a decreasing trend was observed in the biofilm biomass assembled on pre-coated surfaces when compared to the control (RPMI-glucose medium only). On surfaces coated with IHP, 61% (11/18) of ST17 and 65% (13/20) of ST19 strains were classified as NBP (Table 5.4). However, as IHP was added directly with the inoculum the ability of ST17 and ST19 strains to form biofilm increased and 92% of *S. agalactiae* strains (35/38) were classified as SBP for both STs (16 ST17, 19 ST19) (Supplementary material Table S5.1).

Table 5.4 Effect of IHP on biofilm formation by *S. agalactiae* ST17 and ST19 strains.
S. agalactiae biofilm assembly was followed for 48 h on surface conditioned with 15% inactivated human plasma (IHP) for 24 h or not. Biofilms were monitored on non-treated surfaces in RPMI-glucose medium only (control) or supplemented with 15% IHP. For surface conditioned RPMI-glucose medium only was used.

Sequence type	Biofilm assembly category [a]	Isolates (n / %)		
		RPMI-glucose	RPMI-glucose+15%IHP (Pre-conditioning 24 h)	RPMI-glucose+15%IHP
ST17 (n=18)	Non-producer	0	11 (61.1%)	0
	Week	7 (38.9%)	7 (38.9%)	1 (5.6%)
	Moderate	8 (44.4%)	0	1 (5.6%)
	Strong	3 (16.7%)	0	16 (88.9%)
ST 19 (n=20)	Non-producer	0	13 (65%)	0
	Week	2 (10%)	2 (10%)	1 (5%)
	Moderate	16 (80%)	3 (15%)	0
	Strong	2 (10%)	2 (10%)	19 (95%)

[a] Biofilm assembly was classified according to Stepanovic and colleagues [261]

5.3.2. Scanning Electron Microscopy of Biofilms

For *S. agalactiae* reference strains NEM316 and 2603V/R, biofilm assembly was monitored by SEM after 48h incubation in THB supplemented with 1% glucose pH 7.6 and RPMI-glucose (Figure 5.2). For NEM316 a mature biofilm with multiple layers was observed in THB supplemented with 1% glucose pH 7.6 (Figure 5.2A) whereas in RPMI-glucose a thinner phenotype was observed (Figure 5.2B). Nevertheless, for both media, the ability of NEM316 strain to form biofilm was considerably higher than for 2603V/R strain. Independently of using THB supplemented with 1% glucose pH 7.6 (Figure 5.2C) or RPMI-glucose (Figure 5.2D), only a few attached bacterial cells organized in a way compatible with an early-stage biofilm.

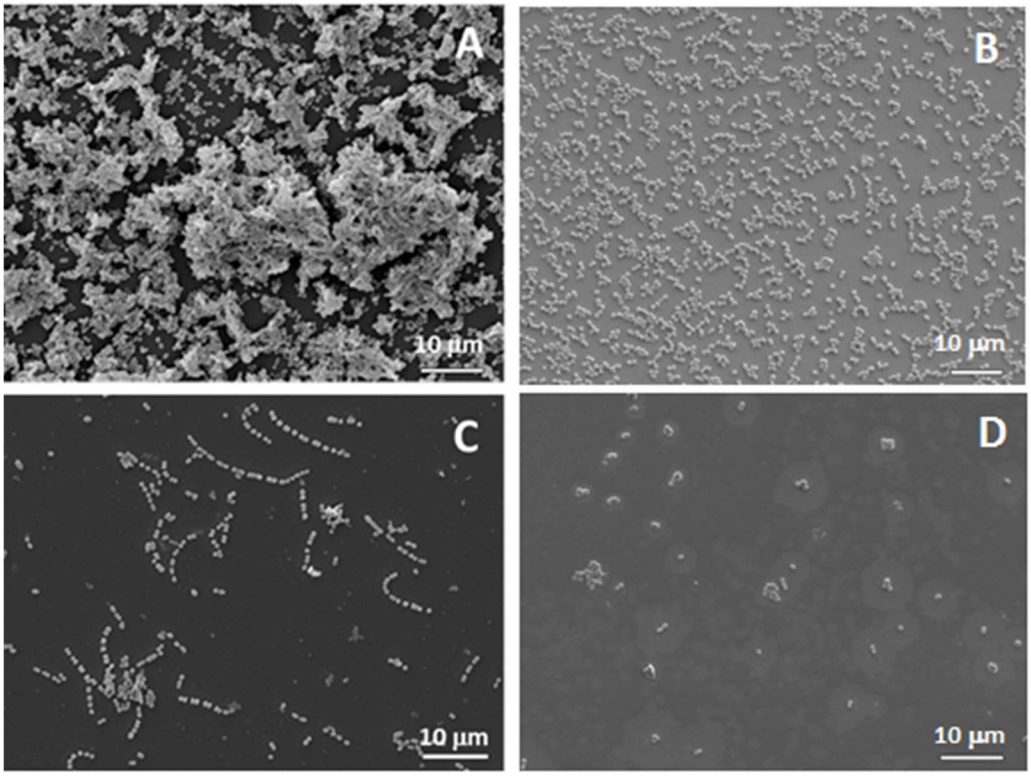


Figure 5.2 *S. agalactiae* biofilms assembled in different conditions. Representative SEM micrographs of biofilm assembly by NEM316 (A-B) and 2603V/R (C-D) on polystyrene incubated at 37ºC/5% CO_2 for 48 h are shown. In the two tested conditions, THB supplemented with 1% glucose at pH 7.6 and in RPMI-glucose, NEM316 was the best biofilm assembler. A well-organized biofilm could be observed in THB supplemented with 1% glucose at pH 7.6 (A) and a less strong phenotype was observed in RPMI-glucose (B) for NEM316. For 2603V/R only an incipient biofilm was detected in both conditions, THB supplemented with 1% glucose at pH 7.6 (C) and RPMI- glucose (D). Scale bar 10 µm.

The CV assay showed a 4-fold increase in OD_{570nm} for NEM316 grown in THB supplemented with 1% glucose pH 7.6, in comparison with 2603V/R. For these conditions, NEM316 was classified as SBP and 2603V/R as WBP. In RPMI-glucose and RPMI-glucose supplemented with 15% IHP, both reference strains were classified as MBP and SBP, respectively. The biofilm formed by NEM316 in RPMI-glucose and RPMI-glucose supplemented with 15% IHP increased only 1.5- and 2-fold, respectively, when compared to 2603V/R.

CV assay results showed that the RPMI-glucose medium attenuated the differences in biofilm formation between the two strains and the addition of IHP induced an alteration in the biofilm phenotype. Comparison of the biofilm assembled by NEM316 in THB supplemented with 1% glucose pH 7.6 (Figure 5.3A) and in RPMI-glucose supplemented with 15% IHP (Figure 5.3B), clearly showed an increase of EPS in the last condition, giving rise to denser biofilms. The presence of EPS forming “bridges” between bacteria could be observed in Figure 5.3B.

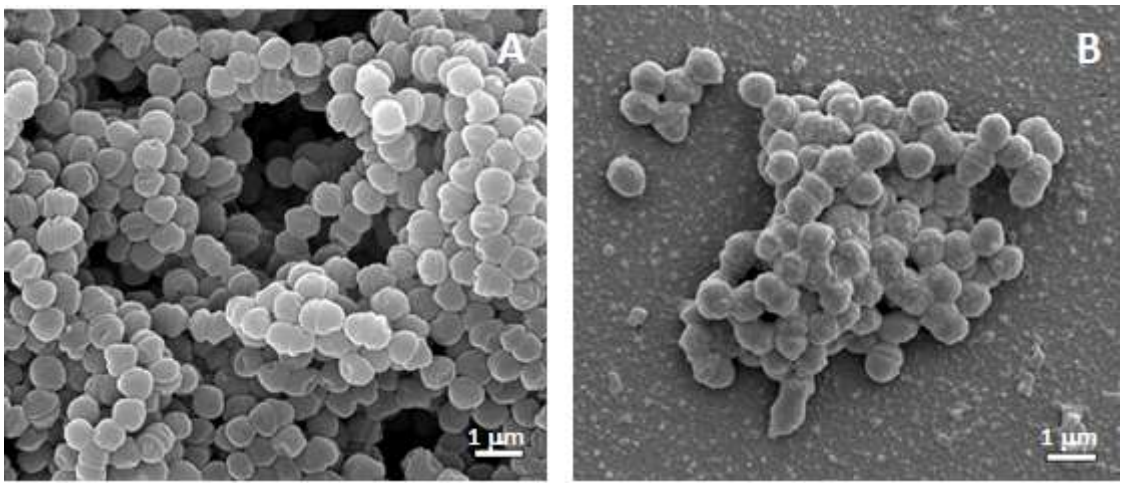


Figure 5.3 Effect of inactivated human plasma on *S. agalactiae* NEM316 biofilms. Biofilm assembly by NEM316 on polystyrene incubated at 37ºC/5% CO_2 for 48 h in THB supplemented with 1% glucose at pH 7.6 (A) and in RPMI-glucose supplemented with 15% IHP (B) were monitored by SEM. Denser biofilms, with increased EPS content, were observed for NEM316 when RPMI-glucose supplemented with 15% IHP (B) was used instead of THB supplemented with 1% glucose at pH 7.6 (A). Scale bar 1 µm.

5.3.3. Correlation of S. agalactiae invasiveness with DNase activity and biofilm formation

A potential relation between the phenotypes of 'strong biofilm formation' and 'virulence' was not observed since no significant difference ($P > 0.05$) could be observed between the biofilm formation capacity of strains from colonizing and infection groups in any of the tested conditions. Additionally, a correlation between DNase activity and biofilm formation was also considered. Significantly higher ($P = 0.035$) biofilm formation was observed in THB supplemented with 1% glucose for DNase producers strains. Notably, in our experiment, ST17 strains (all DNase producers, except 2211-04) seemed to be more prone to assemble biofilms than ST19 strains (all DNase non-producers, except 1203-05). Nonetheless, among the 38 strains, only three (1234-05; 1392-05; 1532-07 - 7.9%; ST17 colonizing strains and DNase producers) could be classified as SBP when incubated in THB supplemented with 1% glucose. With the presence of IHP in the cell culture medium (RPMI-glucose supplemented with 15% IHP) ST19 strains evidenced higher amounts of biofilm when compared with *S. agalactiae* ST17 strains ($P = 0.002$). Nonetheless, in RPMI-glucose medium only (used as control) 80% of ST19 strains (16/20) were classified as MBP.

For *S. agalactiae* reference strain NEM316 (ST23, DNase producer), a 4-fold increase in OD_{570nm} was observed in THB supplemented with 1% glucose at pH 7.6 when compared with 2603V/R (ST110, DNase non-producer).

5.3.4. Composition of S. agalactiae biofilm EPS

Enzymatic dispersion of biofilms is one of the methods used to elucidate the nature of the EPS produced by a specific microorganism [251]. The best conditions to induce biofilm formation (48 h incubation; THB supplemented with 1% glucose; RPMI-glucose supplemented with 15% IHP) and the three ST17 strains (1234-05; 1392-05; 1532-07) classified as SBP in both conditions were selected for evaluation of the EPS composition. To the best of our knowledge, the nature of *S. agalactiae* biofilm EPS is still poorly elucidated, with few reported studies [264]. We performed the enzymatic digestion of 48 h old biofilms by proteinase K, DNase (commercially available DNase I and DNase secreted by NEM316) and β-N-acetylglucosaminidase. Digestion with DNase (DNase I and NEM316 culture supernatant) and β-N-acetilglucosaminidase, did not provide significantly different biofilm dispersions ($P > 0.05$). The major difference in biofilm dispersion was observed for proteinase K digestion (Figure 5.4), inducing a significant difference ($P < 0.05$) in biofilm dispersion. The differences between biofilm dispersion in THB supplemented with 1% glucose and RPMI-glucose supplemented with 15% IHP could be related with the increase of EPS in the last condition. However, these findings suggest that the EPS of *S. agalactiae* biofilms could be richer in proteins than in secreted nucleic acids or exopolysacharides.

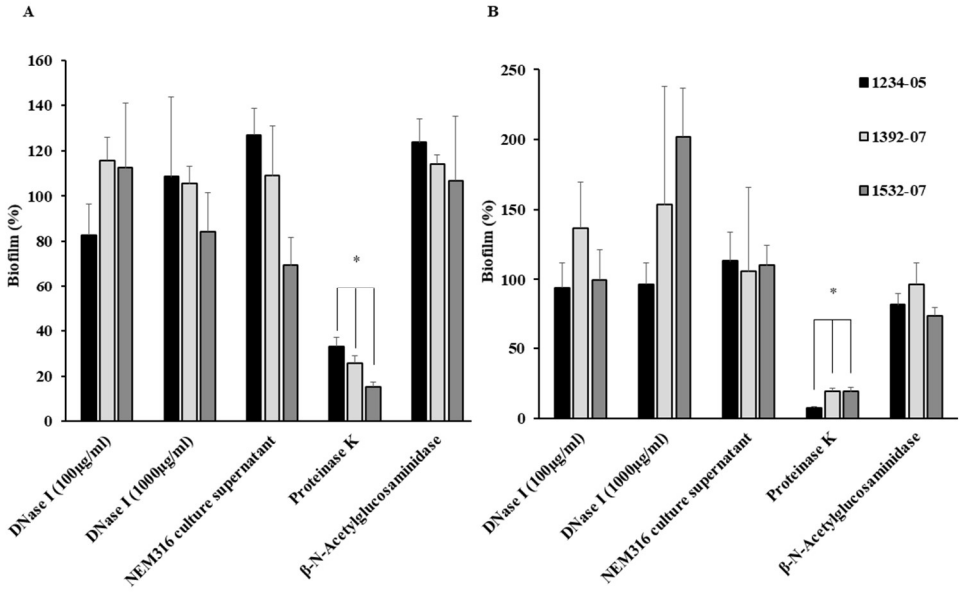


Figure 5.4 The extracellular polymeric substance matrix of *S. agalactiae.* Analysis of the effect of DNase I (100 µg/ml and 1000 µg/ml), culture supernatant from stationary phase of reference strain NEM316, Proteinase K (100 µg/ml) and β-N-acetylglucosaminidase (1.2 units) on 48h-old biofilms in THB supplemented with 1% glucose (A) and in RPMI-glucose supplemented with 15% IHP (B). A significant statistical difference was observed for biofilm dispersion by proteinase K (* $P < 0.05$).

5.4. Discussion

Biofilm studies are impaired by the lack of standardized methods for microorganisms in general [265, 266]. In the case of *S. agalactiae*, an important limitation of most methods applied for biofilm production measurement is that they do not 'exactly' mimic *in vivo* situations, i.e. they lack the dynamic host/pathogen interactions [267]. Moreover, tissue culture-treated plates supplied by different manufacturers may not provide identical conditions for biofilm cultivation and quantification [261]. Thus, the discrepancies in the capacity of *S. agalactiae* to form biofilm-like structures *in vitro* might be explained by the use of different assays and growth conditions. Therefore, one of the aims of the present study was to shed light on the influence of environmental factors (presence *vs* absence of glucose and human plasma, and different pH values) in biofilm formation by *S. agalactiae* human strains of ST17 and ST19, from invasion and from colonization, DNase-producers and non-DNase-producers. The standardization of protocols for biofilm assays and its adoption by the scientific community will allow the comparison of results from different laboratories. Furthermore, the analysis of biofilm assembly in conditions that mimic physiological features will produce results that are more robust and useful to understand bacterial behaviour in vivo [268].

Enhanced biofilm formation was observed for all 38 strains after an incubation of 48 h at 37ºC/5% CO_2 with THB supplemented with 1% glucose at pH 7.6 and with RPMI-glucose supplemented with 15% of IHP. This result is in accordance with the described by other authors [157, 169] who demonstrated that *S. agalactiae* biofilm formation was induced by the presence of 1% glucose in THB. In fact, the composition of the culture medium was considered to be the most important factor to influence the ability of bacteria to produce biofilms, under *in vitro* conditions [261]. The effect of glucose on biofilm formation was proposed to be related with media acidification, as a consequence of the metabolic production of organic acids [169].

Concerning the pH, some authors referred that low pH promotes biofilm production [126, 168, 172]. For *S. agalactiae*, maximum biofilm formation was estimated to occur after a 48 h incubation period [174] and to be influenced by environmental pH changes [234]. The slower bacterial growth at a lower pH was proposed to contribute for the lower biofilm biomass observed at pH 4.5 when compared to the obtained at pH 7.6 [126, 174]. *In vitro* reduced *S. agalactiae* biofilm growth at low pH (< 4.5) has been reported [171, 173, 174]. Indeed, an increase in biofilm formation was observed for strains grown in THB supplemented with 1% glucose at pH 7.6 in comparison to pH 4.5; a similar observation was obtained for pH 7.0. Thus, relatively small pH modifications might be enough to induce important changes in biofilm formation. Yang *et al.* [174] described that *S. agalactiae* survival may be prolonged by growth within a biofilm environment (although reduced at low pH), strengthening the protective effect of biofilms.

Although *S. agalactiae* may adhere to vaginal epithelial cells at low pH, as it is part of the female genital tract commensal flora [269], it might not produce biofilm under vaginal acidic conditions. In fact, *Lactobacillus* species, an important component of the vaginal microbiota, are believed to protect the host from various microbial pathogens by lowering the environmental pH through lactic acid production. In addition, during pregnancy, the species diversity within the vaginal microbiota is reduced but this also drives to further reduction of the vaginal pH; again, low pH seems to have an important protective role against infection for the mother and the fetus [34]. We could hypothesize that upon adverse environmental conditions (such as low pH during pregnancy), *S. agalactiae* would not produce biofilm, but could benefit from biofilms developed by other microbes to colonize the women vaginal tract.

Extracellular nucleases are one of the factors involved in the development of the biofilm architecture [255]. Recently, DNase production was reported for a high percentage of *S. agalactiae* human strains [128]. For this reason, a possible correlation between DNase activity and biofilm formation was evaluated; however, it led to inconclusive results. The best performances for biofilm assembly by ST17 (all DNase producers, except one) and ST19 (all DNase non-producers, except one) *S. agalactiae* strains were observed in different conditions. Stronger biofilm formation was observed in THB supplemented with 1% glucose for ST17 *S. agalactiae* strains. Whereas incubation with RPMI-glucose supplemented with 15% IHP of ST19 *S. agalactiae* strains resulted in higher amounts of biofilm production. The reasons underlying this observation are difficult to determine through the present study; however, they might be related with the capsular polysaccharide, which promotes binding to human plasma proteins, and which is different among ST19 strains (serotypes II, III-1 and V) and ST17 strains (serotype III-2). This possibility should be further evaluated, namely to refine the selection of serotypes to be applied in the development of a vaccine to prevent *S. agalactiae* invasive infection in neonates, elderly and/or immunocompromised people.

In an attempt to correlate *S. agalactiae* invasiveness with biofilm formation, the results from *S. agalactiae* ST17 and ST19 colonization and infection strains were compared. However, no significant differences in the biofilm biomass could be observed. Nonetheless, the reduced number of invasive strains, of both STs, available for the present study could introduce a bias; therefore, it would be of interest to evaluate biofilm formation for a larger group of invasive ST17 and ST19 *S. agalactiae* strains.

Regarding EPS composition, the fact that the addition of DNase (DNase I or NEM316 culture supernatant) to 48h-old biofilms resulted in low biofilm disruption, does not necessarily mean the absence of eDNA in the EPS. In fact, Alvim *et al.* [264] identified eDNA as a common component of *S. agalactiae* biofilms, and eDNA is recognized as a matrix component of most biofilms; however, biofilms might resist to DNase action through processes yet to be identified.

Okshevksy *et al.* [252] suggested that other extracellular matrix components replace or complement eDNA within the mature biofilm, or that eDNA is bound by another component that shields it from enzimatic degradation. For this reason, enzymatic degradation of the EPS should combine the activity of enzymes that target polysaccharides and proteins [252].

The higher amount of biofilm formed by *S. agalactiae* ST17 and ST19 strains in RPMI-glucose supplemented with 15% of IHP (compared with the addition of 15% hiFBS to the medium) suggests that biofilm formation could be promoted by the protein content of human plasma. In 2015, Xia *et al.* [130] reported a major role for *S. agalactiae* capsular polysaccharide in biofilm formation in the presence of human plasma. In fact, ST17 strains were shown to be able to bind to human fibrinogen through a cell wall-anchored glycoprotein (serine-rich repeat, Srr2) [130]. Accordingly, in our study, IHP added directly to the medium promoted the formation of biofilms. Dispersal of *S. agalactiae* mature biofilms by Proteinase K, and not by DNase or β-N-acetylglucosaminidase suggests that proteins could play a major role in biofilm assembly.

In conclusion, biofilm production for 38 *S. agalactiae* strains under the same standardized *in vitro* conditions was evaluated and the composition of the EPS of ST17 *S. agalactiae* strains, DNase producers, was addressed. Proteins seem to be a major component of the *S. agalactiae* EPS matrix. In addition, slightly alkaline pH, glucose and plasma supplementation favour biofilm formation for ST17 (all DNase producers, with exception of one isolate) and ST19 strains (all DNase non producers, with exception of one isolate). No statistical significant correlation between biofilm formation and the strain source (colonization or invasive infection), nor with DNase production, could be established. Further research on *S. agalactiae* biofilms is crucial to develop new curative approaches, namely new enzymatic therapeutics targeting biofilms (avoiding antibiotic overtreatment and resistance), or preventive approaches by the identification of potential targets for new vaccines.

Acknowledgements

The authors wish to thank Barbara Spellerberg for the kind gift of *S. agalactiae* strains from collection of the Institute of Medical Microbiology and Hygiene (IMMH), Ulm University, Germany. We are grateful to João Paulo Gomes and Alexandra Nunes (Instituto Nacional de Saúde Doutor Ricardo Jorge, INSA, IP) for providing β-N-acetylglucosaminidase enzyme and 96-well flat-bottomed polystyrene cell culture plates, respectively. We thank Rita Sobral (Nova School of Science and Technology, FCT NOVA) for the critical review of the manuscript and Hemovida® for IHP donation.

Chapter VI

Final overview, concluding remarks and future perspectives

6. Final overview, concluding remarks and future perspectives

S. agalactiae is commonly isolated from the lower gastrointestinal tract or the lower female reproductive tract. Though normally an asymptomatic colonizer in these environments, this opportunistic pathogen can cause invasive disease, including pneumonia, sepsis, and meningitis. In the 1960's, this bacterial species emerged as a major cause of neonatal morbidity and mortality, and since the 1990's it has also been highlighted as an agent of invasive infections in non-pregnant adults. In order to establish colonization and/or infection, *S. agalactiae* is able to produce several virulence factors, which contribute to host cell attachment, invasion, colonization and progression to invasive disease [26, 37, 145, 264].

The CPS represents one of the most critical virulence factors in *S. agalactiae*. In particular, III/ST17 strains have been considered to constitute an hypervirulent clone commonly associated with invasive disease, specifically meningitis, while strains from serotype III/ST19 have been associated with colonization [37, 88, 145]. In Chapter II the main virulence factors and genome characteristics of *S. agalactiae* that contribute to the epidemiological success of some clones, particularly regarding the major pathogenic clone III/ST17, were revised. Although the clone III/ST17 is characterized by a specific combination of genetic markers, suggesting that it is a stable clone, the acquisition of MGEs encoding antimicrobial resistance has been reported. In fact, horizontal gene transfer should play a significant role in *S. agalactiae* genome evolution, contributing to diversity among *S. agalactiae* strains and to the emergence of new hypervirulent clones, namely III/ST17 sublineages [32, 132, 270]. *S. agalactiae* colonization and infection depend largely on the bacterial ability to adhere and persist on mucosal surfaces, where virulence factors such as biofilms could facilitate microbial survival and proliferation. The ability to produce biofilm has been highlighted as an important virulence trait among *S. agalactiae* [156, 157, 168, 173]. The development of the three-dimensional biofilm structure would be controlled by extracellular DNases, which were proposed to be involved in the detachment from the mature biofilm and in the utilization of eDNA as a nutrient source [255]. Furthermore, extracellular DNases should have a major role in bacterial adaptation to the host immune response, contributing to the evasion from NET-mediated antimicrobial activity.

Nuclease activity in *S. agalactiae* was demonstrated in 1980s by Ferrieri *et al.* [162], and in 2013 the analysis of the genome of *S. agalactiae* reference strain NEM316 revealed the existence of a major nuclease, NucA [163]. The existence of other nucleases in *S. agalactiae* was supported by the residual DNase activity displayed by a *nucA* mutant and was attributed to one or more of the other six putative DNases encoding genes [163]. However, the genomic basis underlying DNase activity in *S. agalactiae* is far from being deciphered, and the exact role of DNases during pathogenesis remains to be explored.

In Chapter III, a comparative genomic analysis by WGS was performed for a collection of ST17 and ST19 human strains, from carriage and infection, presenting different DNase production phenotypes. The phenotypic heterogeneity of strains with the same ST was the basis to evaluate the background of DNase activity. Thus, a deep characterization of the genetic variability among the 'exception strains' (ST17 DNase(-) and ST19 DNase(+)) was performed. In this regard, Chapter III adds knowledge on *S. agalactiae* ST17 and ST19 genomes through full sequencing and public release of the sequences. The *nucA* gene and the other six putative DNase encoding genes were identified for all strains. Specific alleles were exhibited for *nucA* by the sole ST17 DNase(-) strain, and for one of the other putative DNase genes (*gbs0609*) by the sole ST19 DNase(+) strain. Although it is very likely that, besides NucA and other predicted DNases, other proteins, namely those involved in their export/secretion, may account for the DNase phenotypic differences observed between ST17 and ST19 strains, we speculate that these mutations could sustain the distinct DNase phenotype of these two strains. A global examination of the core-genome allowed to identify core-SNVs or indels that potentially contribute to DNase activity. However, the impact of these mutations in DNase enzymatic performance remains undetermined. Due to its genomic diversity, *S. agalactiae* was considered an *open pan-genome* [11, 111, 112]; as such we aimed to identify accessory genes that could discriminate the strains exhibiting a different DNase production phenotype. Many of these acquired specific genes have been associated with mobile genetic elements. In fact, the ST17 DNase(-) strain was found to harbor a unique intact phage with no homology in the available *S. agalactiae* genome sequences, and a transposon homologous to TnGBS2.3. Although this transposon is shared with another ST17 strain (DNase(+) phenotype), the site of insertion in the DNase(-) strain is located near genes that belong to a specialized secretion system (type VII protein secretion system), responsible for transport of proteins across the cell membrane to the extracellular environment. This type VII protein secretion system has been associated with virulence in *Staphylococcus aureus* [232]. The impact of the transposon insertion in the expression of this system and the link to the DNase non-producer phenotype could not be determined through the present study, but it surely deserves further evaluation. The identification of nine accessory genes among all ST17 DNase(+) strains and the observation that none of these loci was present in any of the DNase(-) strains may suggest a contribution of these proteins to the lack of DNase activity. Nevertheless, this hypothesis also warrants further investigation through functional assays. Overall, we globally characterized the genome of *S. agalactiae* strains from ST17 and ST19 with different DNase phenotypes in an attempt to shed some light on the genetic determinants of *S. agalactiae* associated to DNase activity. Despite the need to confirm these data with other *S. agalactiae* strains of different ST and diverse DNase production phenotypes, new data was revealed regarding putative targets to be used in future functional studies aiming to clarify the DNase activity.

In Chapter IV we performed RNA-seq preliminary assays for evaluating *S. agalactiae* gene expression at exponential growth phase and whether it is affected by the presence of human DNA exposure (mimicking NETs). However, no differentially expressed gene was detected for NEM316 at the exponential phase as a consequence of the presence of human DNA. Indeed, very little is known about the complex molecular cascades mediating DNase release *in vitro* and *in vivo* and streptococcal DNases have been shown to be under the control of extensive regulatory systems [138]. Thus, it would be interesting to test other stimuli and perform comparative RNA-seq assays involving both high and low DNase producing strains in comparison with DNase non-producers. Our initial hypothesis for Chapter IV was not confirmed; nevertheless, the global gene expression dynamics of NEM316 *S. agalactiae* strain during the exponential phase could be evaluated. As a means to provide the more complete data required to better interpret and discuss results, firstly we analyzed the genome backbone of the reference strain NEM316 subjected to RNA-seq, to evaluate whether laboratory passaging could induce significant genetic changes when comparing with the publicly available genome. Subsequently, genes were ranked by mean expression level and grouped by functional category. The exponential phase was chosen because it is particularly interesting when studying the transcriptional activity as it ensures that the expression of most of the genes is assessed. The functional category "Translation, ribosomal structure and biogenesis" represented 46% of the top-100 expressed genes, which is not surprising as a high demand for proteins related to translation and metabolism should occur during the exponential growth phase. Interestingly, genes not assigned to any functional category were also represented among the highly expressed genes. Although most of them seem to be implicated in the bacterial stress response, many do not have any assigned function, neither any predicted functional domain. We were also interested in the analysis of the mean expression levels of genes related to virulence, including those that encode for DNases, PI (involved in adhesion and biofilm formation) and other genes included in the virulome characterization described in Chapter III. Curiously, *nucA* and PI-2a genes presented higher mean expression values than the other putative DNase and PI-1 genes, respectively. The PI-2a pilus major subunit PilB was ranked in the top-100 most expressed genes, as well as two genes coding for adhesins. Globally, the results of this Chapter represent a first attempt to systematize the expression levels of *S. agalactiae* reference strain NEM316 at exponential phase during their normal growth in laboratory conditions. Certainly, it constitutes an important database for further studies.

In Chapter V the effect of environmental factors (H^+, glucose and human plasma) in biofilm formation was evaluated *in vitro*, under standardized experimental conditions, for ST17 and ST19 *S. agalactiae* strains. SEM technique was used to evaluate the biofilm phenotype. We also accessed the composition of the extracellular polymeric matrix of the three strongest biofilm producers (all from ST17) by enzymatic digestion of mature biofilms. As biofilms should play an

important role in *S. agalactiae* pathogenicity and virulence, the conflicting data reported, regarding the biofilm-forming capacity for *S. agalactiae* strains, justified the need for standardization of experimental procedures, which will allow the comparison of results obtained in different laboratories. In addition, we aimed to correlate DNase activity and biofilm formation; however, no statistical significant correlation could be established. The performances of biofilm assembly by ST17 and ST19 *S. agalactiae* strains were shown to depend on the conditions tested, which may be related with the specificities of the capsular polysaccharide, involved in biofilm assembly. The composition of the culture medium was considered the most important factor influencing the ability of bacteria to produce biofilms under *in vitro* conditions. In fact, glucose, plasma and pH 7.6 were shown to enhance biofilm formation in the present study. Considering the pH, minor pH modifications might be enough to induce important changes in biofilm formation. Regarding the EPS composition, proteins were shown to be the predominant component of the *S. agalactiae* matrix. These results should contribute to the development of new enzymatic therapeutics targeting biofilms during *S. agalactiae* infections. Overall, optimal conditions for *S. agalactiae* biofilm assembly were determined through this study, contributing for the standardization of protocols for biofilm assays.

In conclusion, the findings presented in this Ph.D. thesis should contribute for the understanding of the biological role of DNases in *S. agalactiae*, and provide new insights into biofilm formation, genomics and transcriptomics for ST17 and ST19 lineages. The perfect knowledge of the determinants for DNase production and the ecological success dissimilarities among *S. agalactiae* strains remains to be understood but may involve genome structures highlighted throughout this Ph.D. dissertation.

Several interesting observations arose during the course of this Ph.D. thesis which we believe should be further investigated in the near future; specific lines of work would include:

- Clarifying the genetic background of DNases and their biological role integrating genomics and transcriptomics data with proteomics, metabolomics and epigenomics. As the kinetic measurements should provide a complete and more realistic picture of the dynamics of nucleases' catalytic activity, we established a collaboration with a Sweden group for kinetic screening of DNases using nucleic acid probes [271]. This study is ongoing at the time this thesis was completed;

- Understanding the biological role of the genes not assigned to any functional category that presented intriguing expression values being among the top-100 most expressed genes for *S. agalactiae* NEM316 at exponential phase (Chapter IV);

- Identifying the molecular determinants underlying ST17 and ST19 best performances for biofilm assembly in different conditions. Investigating the role played by the capsular polysaccharide in biofilm assembly.

References

1. **Nobbs, A.H., R. J. Lamont, and H. F. Jenkinson.** 2009. *Streptococcus* adherence and colonization. Microbiol Mol Biol Rev. **3**:407-450.
2. **Facklam, R.** 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin Microbiol Rev. **15**:613-630.
3. **McPherson, R. A. and M. R. Pincus.** 2007. Henry's Clinical Diagnosis and Management by Laboratory Methods (21 ed.). Saunders, Philadelphia.
4. **Ruoff, K. L., R. A. Whiley, and D. Beighton.** 2003. *Streptococcus*, p.405-417. *In* P. R. Murray, E. J. Baron, J. H. Jorgensen, M. A. Pfaller, and R. H. Yolken (8 ed.), Manual of Clinical Microbiology, vol. 1. ASM Press, Washington, DC.
5. **Motarjemi, Y.** 2014. Encyclopedia of Food Safety (2nd ed.). Elsevier, Oxford.
6. **Lancefield, R.C.** 1934. A serological differentiation of specific types of bovine hemolytic streptococci (Group B). J Exp Med. 59(4):441-458.
7. **Quah, S. R. and W. Cockerham.** 2017. International Encyclopedia of Public Health (2nd ed.). Elsevier, Oxford.
8. **Tomida, J., S. Sauer, Y. Morita, T. Ezaki, and Y. Kawamura.** 2011. *Streptococcus fryi* sp. nov., a novel species with Lancefield group M antigens. FEMS Microbiol Lett. **314**:95-100.
9. **Rosa-Fraile, M., and B. Spellerberg.** 2017. Reliable detection of group B Streptococcus in the Clinical Laboratory. J Clin Microbiol. **55**(9):2590-2598.
10. **Gao, X., Zhi X., Li H., Klenk H., and Li W.** 2014. Comparative Genomics of the Bacterial Genus *Streptococcus* Illuminates Evolutionary Implications of Species Groups. PLoS One. **9**(6):e101229.
11. **Glaser, P., C. Rusniok, C. Buchrieser, F. Chevalier, L. Frangeul, T. Msadek, M. Zouine, E. Couvé, 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.
12. **Ke, D., C. Menard, F. J. Picard, M. Boissinot, M. Ouellette, P. H. Roy, M. G. Bergeron.** 2000. Development of conventional and real-time PCR assays for the rapid detection of group B streptococci. Clin Chem. **46**:324-331.
13. **Lancefield, R.C., and R. Hare.** 1935. The serological differentiation of pathogenic and non-pathogenic strains of hemolytic *Streptococci* from parturient women. J Exp Med. **61**:335-349.
14. **Fry, R. M.** 1938. Fatal infections by haemolytic streptococcus group B. Lancet. **1**:199-200.
15. **Beitune, P., G. Duarte., and C. M. L. Maffei.** 2005. Colonization by *Streptococcus agalactiae* During Pregnancy: Maternal and Perinatal Prognosis. The Brazilian Jounal of Infectious Diseases. **9**:276-282.
16. **Koening, J.M., and W. J. Keenan.** 2009. Group B *Strepotococcus* and Early-Onset Sepsis in the Era of Maternal Prophylaxis. Pediatr Clin North Am. **56**:689-710.
17. **Eickhoff, T. C., J. O. Klein, A. K. Daly, D. Ingall, M. Finland.** 1964. Neonatal Sepsis and Other Infections Due to Group B Beta-Hemolytic Streptococci. The New England Journal of Medicine. **271**(24):1221-1128.
18. **Barton, L. L., R. D. Feigin, and R. Linds.** 1973. Group B beta hemolytic streptococcal meningitis in infants. The Journal of Pediatrics. **82**(4):719-723.
19. **Baker, C. J., F. F. Barrett, R. C. Gordon, and M. D. Yow.** 1973. Suppurative meningitis due to streptococci of Lancefield group B: A study of 33 infants. The Journal of Pediatrics. **82**(4):724-729.
20. **Franciosi, R. A., J. D. Knostman, and R. A. Zimmerman.** 1973. Group B streptococcal neonatal and infant infections. The Journal of Pediatrics. **82**(4):707-718.
21. **Howard, J. B., and G. H. McCracken.** 1974. The Spectrum of Group B Streptococcal Infections in Infancy. Am J Dis Child. **128**:815-818.
22. **Centers for Disease Control and Prevention.** 2010. Prevention of Perinatal Group B Streptococcal Disease: Revised Guidelines. MMWR. **59**:1-23.

23. **Gibbs, R.S., S. Schrag, and A. Schuchat.** 2004. Perinatal Infections Due to Group B *Streptococci*. Obstetrics & Gynecology. **104**:1062-1076.
24. **Melin, P.** 2011. Neonatal group B *Streptococcal* disease: from pathogenesis to preventive strategies. Clin Microbiol Infect. **17**:1294-1303.
25. **Patterson, M. J., and A. B. Hafeez.** 1976. Group B Streptococci in Human Disease. Bact Rev. **40**:774-792.
26. **Edmond, K. M., C. Kortsalioudaki, S. Scott, S. J. Schrag, A. K. Zaidi, S. Cousens, P. T. Heath.** 2012. Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. Lancet. **379**:547–556.
27. **Dzanibe, S., and S. A. Madhi.** 2018. Systematic review of the clinical development of Group B Streptococcus serotype specific capsular polysaccharide based vacines. Expert Review of Vaccines. **17**:7.
28. **Seale, A. C., C. J. Baker, J. A. Berkley, S. A. Madhi, J. Ordi, S. K. Saha, S. J. Schrag, A. S. Meulen, J. Vekemans.** 2019. Vaccines for maternal immunization against Group B Streptococcus disease: WHO perspectives on case ascertainment and case definitions. Vaccine. **37**:4877-4885.
29. **Honest, H., S. Sharma, and K. S. Khan.** 2006. Rapid Tests for Group B *Streptococcus* Colonization in Laboring Women: A Systematic review. Pediatrics. **117**:1055-1066.
30. **Schrag, S.J., S. Zywicki, M. M. Farley, A. L. Reingold, L. H. Harrison, L. B. Lefkowitz, J. L. Hadler, R. Danila, P. R. Cieslak, and A. Schuchat.** 2000. Group B *Streptococcal* Disease in The Era of Intrapartum Antibiotic Prophylaxis. The New England Journal of Medicine. **342**:15-20.
31. **Schuchat, A.** 1998. Epidemiology of group B *Streptococcal* disease in the United States: shifting paradigms. Clin Microbiol Rev. **11**:497-513.
32. **Martins, E. R., C. Pedroso-Roussado, J. Melo-Cristino, M. Ramirez, and The Portuguese Group for the Study of Streptococcal Infections.** 2017. *Streptococcus agalactiae* Causing Neonatal Infections in Portugal (2005-2015): Diversification and Emergence of a CC17/PI-2b Multidrug Resistant Sublineage. Front Microbiol. **8**:499.
33. **Lindahl, G., M. Stalhammar-Carlemalm, and T. Areschoug.** 2005. Surface Proteins of *Streptococccus agalactiae* and Related Proteins in Other Bacterial Pathogens. Clin Microbiol Rev. **18**:102-127.
34. **Patras, K. A., and V. Nizet.** 2018. Group B Streptococcal Maternal Colonization and Neonatal Disease: Molecular Mechanisms and Preventative Approaches. Front Pediat. **6**:27.
35. **Raabe, V. N., and A. L. Shane.** 2019. Group B Streptococcus (*Streptococcus agalactiae*). Microbiol Spectr. **7**:2.
36. **Brown, A. P., and F. C. Denison.** 2018. Selective or universal screening for GBS in pregnancy (review). Early Human Development. **126**:18-22.
37. **Shabayek, S., and B. Spellerberg.** 2018. Group B Streptococcal Colonization, Molecular Characteristics, and Epidemiology. Front Microbiol. **9**:437.
38. **ACOG Committee Opinion.** 2020. Prevention of Group B Streptococcal Early-Onset Disease in Newborns. Obstetrics & Gynecology. **135**(2):e51-72.
39. **Jahromi, B. N., S. Poorarian, and S. Poorbarfehee.** 2008. The Prevalence and Adverse Effects of Group B *Streptococcal* Colonization during Pregnancy. Archives of Iranian Medicine. **11**:654-657.
40. **Manning, S. D., K. Neighbors, P. A. Tallman, B. Gillespie, C. F. Marrs, S. M. Borchardt, C. J. Baker, M. D. Pearlman, and B. Foxman.** 2004. Prevalence of Group B *Streptococcus* Colonization and Potential for Transmission by Casual Contact in Healthy Young Men and Women. Clin Infect Dis. **39**:380-388.
41. **Daniels, J. P., J. Gray, H. M. Pattinson, R. Gray, R. K. Hills, and K. S. Khan on behalf of the GBS Collaborative Group.** 2011. Intrapartum tests for group B streptococcus: accuracy and acceptability of screening. BJOG. **118**:257-265.
42. **Florindo C., V. Damião, J. Lima, I. Nogueira, I. Rocha, P. Caetano, L. Ribeiro, S. Viegas, J. P. Gomes, and M. J. Borrego.** 2014. Accuracy of prenatal culture in predicting

intrapartum group B streptococcus colonization status. J Matern Fetal Neonatal Med. **27**(6): 640–642.
43. **Neto, M. T.** 2008. Group B Streptococcal Disease in Portuguese Infants Younger than 90 Days. Arch Dis Child Fetal Neonatal Ed. **93**(2):F90-3.
44. **Poyart, C., H. Réglier-Poupet, A. Tazi, A. Billoet, N. Dmytruk, P. Bidet, E. Bingen, J. Raymond, and P. Trieu-Cuot.** 2008. Invasive Group B Streptococcal Infections in Infants, France. Emerg Infect Dis. **14**:1647-1649.
45. **Farley, M.** 2001. Group B *Streptococcal* Disease in Nonpregnant Adults. Clin Infect Dis. **33**:556-561.
46. **Kimura, K., S., Suzuki, J-i Wachino, H. Kurokawa, K. Yaamane, N. Shinata, N. Nagano, H. Kato, K. Shibayama, and Y. Arakawa.** 2008. First molecular characterization of Group B *Streptococci* with reduced Penicillin Susceptibility. Antimicrobial Agents and Chemotherapy. **52**:2890-2897.
47. **Gygax, S. E., J. A. Schuyler, L. E. Kimmel, J. P. Trama, E. Mordechai, and M. E. Adelson.** 2006. Erythromycin and clindamycin resistance in group B *Streptococcal* clinical isolates. Antimicrobial Agents and Chemotherapy. **50**:1875-1877.
48. **Fitoussi, F., C. Loukil, I, Gros, O. Clermont, P. Mariani, S. Bonacorsi, I. Le Thomas, D. Deforche, and E. Bingen.** 2001. Mechanisms of macrolide resistance in clinical group B *Streptococci* isolated in France. Antimicrobial Agents and Chemotherapy. **45**:1889-1891.
49. **Tenover, F. C.** 2006. Mechanisms of Antimicrobial Resistance in Bacteria. American Journal of Medicine. **119**:S3-S10.
50. **Weisblum, B.** 1995. Erythromycin Resistance by Ribosome Modification. Antimicrobial Agents and Chemotherapy. **39**:577-585.
51. **Hawkins, P. A., C. S. Law, B. J. Metcalf, S. Chochua, D. M. Jackson, L. F. Westblade, R. Jerris, B. W. Beall, and L. McGee.** 2017. Cross-resistance to lincosamides, streptogramins A and pleuromutilins in *Streptococcus agalactiae* isolates from the USA. J Antimicrob Chemother. **72**:1886-1892.
52. **Dahesh, S., M. E. Hensler, N. M. Van Sorge, R. E. Gertz, Jr., S. Schrag, V. Nizet, and B. W. Beall.** 2008. Point mutation in the group B streptococcal pbp2x gene conferring decreased susceptibility to beta-lactam antibiotics. Antimicrobial Agents Chemotherapy. **52**:2915-2918.
53. **Kimura, K., K. Matsubara, G. Yamamoto, K. Shibayama, and Y. Arakawa.** 2013. Active Screening of Group B Streptococci with Reduced Penicillin Susceptibility and Altered Serotype Distribution Isolated from Pregnant Women in Kobe, Japan. Jpn J Infect Dis. **66**:158-160.
54. **Seki, T., K. Kimura, M. E. Reid, A. Miyazaki, H. Banno, W. Jin, J. Wachino, K. Yamada, and Y. Arakawa.** 2015. High isolation rate of MDR group B streptococci with reduced penicillin susceptibility in Japan. J Antimicrob Chemoter. **70**:2725-2728.
55. **Lopes, E., T. Fernandes, M. P. Machado, J. A. Carriço, J. Melo-Cristino, M. Ramirez, E. R. Martins, and the Portuguese Group for the Study of Streptococcal Infections.** 2018. Increasing macrolide resistance among *Streptococcus agalactiae* causing invasive disease in non-pregnant adults was driven by a single capsular-transformed lineage, Portugal, 2009 to 2015. Euro Surveill. **23**(21):pii=1700473.
56. **Domelier, A-S, N. Mee-Marquet, L. Arnault, L. Mereghetti, P. Lanotte, A. Rosenau, M-F. Lartigue, and R. Quentin.** 2008. Molecular characterization of erythromycin-resistant *Streptococcus agalactiae* strains. J Antimicrob Chemoter. **62**:1227-1233.
57. **Gherardi, G., M. Imperi, L. Baldassarri, M. Pataracchia, G. Alfarone, S. Recchia, G. Orefici, G. Dicuonzo, and R. Creti.** 2007. Molecular epidemiology and distribution of serotypes, surface proteins, and antibiotic resistance among group B *Streptococci* in Italy. J Clin Microbiol. **45**:2909-2916.
58. **Rosato, A., H. Vicarini, and R. Leclercq.** 1999. Inducible or constitutive expression of resistance in clinical isolates of *Streptococci* and *Enterococci* cross-resistant to erythromycin and lincomycin. J Antimicrob Chemoter. **43**:559-562.

59. **De Azavedo, J.C., M. McGavin, C. Duncan, D. E. Low, and A. McGeer.** 2001. Prevalence and mechanisms of macrolide resistance in invasive and noninvasive group B S*treptococcus* isolates from Ontario, Canada. Antimicrobial Agents and Chemotherapy. **45**:3504-3508.
60. **Seppala, H., M. Skurnik, H. Soini, M. C. Roberts, and P. Huovinen.** 1998. A novel erythromycin resistance methylase gene (*ermTR*) in *Streptococcus pyogenes*. Antimicrobial Agents and Chemotherapy. **42**:257-262.
61. **Varaldo, P. E., M. P. Montanari, and E. Giovanetti.** 2009. Genetic Elements Responsible for Erythromycin Resistance in Streptococci. Antimicrobial Agents and Chemotherapy. **53**(2):343-353.
62. **Park, C., M. Nichols, and S. J. Schrag.** 2014. Two Cases of Invasive Vancomycin-Resistant Group B Streptococcus Infection. The New England Journal of Medicine. **370**(9):885-886.
63. **Srinivasan, V., B. J. Metcalf, K. M. Knipe, M. Ouattara, L. McGee, P. L. Shewmaker, A. Glennen, M. Nichols, C. Harris, M. Brimmage, B. Ostrowsky, C. J. Park, S. J. Schrag, M. A. Frace, S. A. Sammons, and B. Beall.** 2014. *vanG* Element Insertions within a Conserved Chromosomal Site Conferring Vancomycin Resistance to *Streptococcus agalactiae* and *Streptococcus anginosus*. mBio. **5**(4):e01386-14.
64. **Hays, C., M. Louis, C. Plainvert, N. Dmytruk, G. Touak, P. Trieu-Cuot, C. Poyart, and A. Tazi.** 2016. Changing Epidemiology of Group B Streptococcus Susceptibility to Fluoroquinolones and Aminoglycosides in France. Antimicrobial Agents and Chemotherapy. **60**(12):7424-7430.
65. **Kawamura, Y., H. Fujiwara, N. Mishima, Y. Tanaka, A. Tanimoto, S. Ikawa, Y. Itoh, and T. Ezaki.** 2003. First *Streptococcus agalactiae* Isolates Highly Resistant to Quinolones, with Point Mutations in *gyrA* and *parC*. Antimicrobial Agents and Chemotherapy. **47**:3605-3609.
66. **Ki, M., U. Srinivasan, K. Y. Oh, M. Y. Kim, J.-H. Shin, H. L. Hong, T. Dang, Z. Britt, and B. Foxman.** 2012. Emerging fluoroquinolone resistance in *Streptococcus agalactiae* in South Korea. Eur J Clin Microbiol Infect Dis. **23**:1685-1691.
67. **Wehbeh, W., R. Rojas-Diaz, X. Li, N. Mariano, L. Grenner, S. Segal-Maurer, B. Tommasulo, K. Drlica, C. Urban, and J. J. Rahal.** 2005. Fluoroquinolone-Resistant *Streptococcus agalactiae*: Epidemiology and Mechanism of Resistance. Antimicrobial Agents and Chemotherapy. **49**:2495-2497.
68. **Teatero, S., E. Ramoutar, A. McGeer, A. Li, R. G. Melano, J., K. Dewar, and N. Fittipaldi.** 2016. Clonal Complex 17 Group B *Streptococcus* strains causing invasive disease in neonates and adults originate from the same genetic pool. Scientific Reports. **6**:20047.
69. **Da Cunha, V., M. R. Davies, P. E. Douarre, I. Rosinski-Chupin, I. Margarit, S. Spinali, T. Perkins, P. Lechat, N. Dmytruk, E. Sauvage, L. Ma, B. Romi, M. Tichit, M. J. Lopez-Sanchez, S. Descorps-Declere, E. Souche, C. Buchrieser, P. Trieu-Cuot, I. Moszer, D. Clermont, D. Maione, C. Bouchier, D. J. McMillan, J. Parkhill, J. L. Telford, G. Dougan, M. J. Walker, The DEVANI Consortium, M. T. G. Holden, C. Poyart, and P. Glaser.** 2014. *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. Nat Commun. **5**:4544.
70. **Chen, S. L.** 2019. Genomic Insights Into the Distribution and Evolution of Group B Streptococcus. Front Microbiol. **10**:1447.
71. **Emaneini, M.**, **A. Mirsalehian, R. Beigvierdi, A. A. I. Fooladi, F. Asadi, F. Jabalameli, M. Taherikalani.** 2014. High Incidence of Macrolide and Tetracycline Resistance among *Streptococcus agalactiae* Strains Isolated from Clinical Samples in Tehran, Iran. Maedica. 9(2):157-161.
72. **Ramaswamy, S.V., P. Ferrieri, A. E. Flores, and L. C. Paoletti.** 2006. Molecular Characterization of Nontypable Group B *Streptococcus*. J Clin Microbiol**. 44**:2398-2403.
73. **Ramaswamy, S.V., P. Ferrieri, L. C. Madoff, A. E. Flores, N. Kumar, H. Tettelin, and L. C. Paoletti.** 2006. Identification of a novel *cps* locus polymorphisms in nontypable group B *Streptococcus.* J Med Microbiol. **55**:775-783.

74. **Bohnsack, J.F., A. Whiting, M. Gottschalk, D. M. Dunn, R. Weiss, P. H. Azimi, J. B. Philips, L. E. Weisman, G. G. Rhoads, and F. Y. Lin.** 2008. Population structure of invasive and colonizing strains of *Streptococcus agalactiae* from neonates of six U.S. Academic Centers from 1995 to 1999. J Clin Microbiol. **46**:1285-1291.
75. **Quentin, R., H. Huet, F. S. Wang, P. Geslin, A. Goudeau, and R. K. Selander.** 1995. Characterization of *Streptococcus agalactiae* strains by multilocus enzyme genotype and serotype: identification of multiple virulent clone families that cause invasive neonatal disease. J Clin Microbiol. **33**:2576-2581.
76. **Slotved, H.C., F. Kong, L. Lambertsen, S. Sauer, and G. L. Gilbert.** 2007. Serotype IX, a Proposed New *Streptococcus agalactiae* Serotype. J Clin Microbiol. **45**:2929-2936.
77. **Breeding, K. M., B. Ragipani, K. D. Lee, M. Malik, T. M. Randis, A. J. Ratner.** 2016. Real-time PCR-based serotyping of *Streptococcus agalactiae*. Nature. **6**:38523.
78. **Kapatai, G., D. Patel, A. Efstratiou, V. J. Chalker.** 2017. Comparison of molecular serotyping approaches of *Streptococcus agalactiae* from genomic sequences. BMC Genomics. **18**:429.
79. **Kong, F., L. M. Lambertsen, H-C. Slotved, D. Ko, H. Wang, and G. L. Gilbert.** 2008. Use of Phenotypic and Molecular Serotype Identification Methods To Characterize Previously Nonserotypeable Group B *Streptococci*. J Clin Microbiol. **46**:2745-2750.
80. **Palacios, G. C., E. K. Eskew, F. Solorzano, and S. J. Mattingly.** 1997. Decreased capacity for type-specific-antigen synthesis accounts for high prevalence of nontypeable strains of group B *Streptococci* in Mexico. J Clin Microbiol. **35**:2923-2926.
81. **Sellin, M., C. Olofsson, S. Hakansson, and M. Norgren**. 2000. Genotyping of the capsule gene cluster (*cps*) in nontypeable group B *Streptococci* reveals two major *cps* allelic variants of serotypes III and VII. J Clin Microbiol. **38**:3420-3428.
82. **Kong, F., S. Gowan, D. Martin, G. James, and G. L. Gilbert.** 2002. Serotype Identification of Group B *Streptococci* by PCR and Sequencing. J Clin Microbiol. **40**:216-226.
83. **Cieslewicz, M. J., D. Chaffin, G. Glusman, D. Kasper, A. Madan, S. Rodrigues, J. Fahey, M. R. Wessels, and C. E. Rubens.** 2005. Structural and Genetic Diversity of Group B *Streptococcus* Capsular Polysaccharides. Infection and Immunity. **73**:3096-3103.
84. **Manning, S. D., D. W. Lacher, H. D. Davies, B. Foxman, and T. S. Whittam.** 2005. DNA Polymorphism and Molecular Subtyping of the Capsular Gene Cluster of Group B *Streptococcus*. J Clin Microbiol. **43**:6113-6116.
85. **Luan, S., M. Granlund, M. Sellin, T. Lagergård, B. G. Spratt, and M. Norgren.** 2005. Multilocus Sequence Typing of Swedish Invasive Group B *Streptococcus* Isolates Indicates a Neonatally Associated Genetic Lineage and Capsule Switching. J Clin Microbiol. **43**:3727-3733.
86. **Lachenauer, C. S., D. L. Kasper, J. Shimada, Y. Ichiman, H. Ohtsuka, M. Kaku, L. C. Paoletti, P. Ferrieri, and L. C. Madoff.** 1999. Serotypes VI and VIII Predominate among Group B Streptococci Isolated from Pregnant Japanese Women. J Infect Dis. **179**:1030-1033.
87. **Florindo C., V. Damião, I. Silvestre, C. Farinha, F. Rodrigues, F. Nogueira F, F. Martins-Pereira, R. Castro, M. J. Borrego, I. Santos-Sanches, and The Group for the Prevention of Neonatal GBS Infection.** 2014. Epidemiological surveillance of colonising group B *Streptococcus* epidemiology in the Lisbon and Tagus Valley regions, Portugal (2005 to 2012): emergence of a new epidemic type IV/clonal complex 17 clone. Euro Surveill. **19**:23.
88. **Martins, E. R., M. A. Pessanha, M. Ramirez, and J. Melo-Cristino.** 2007. Analysis of Group B Streptococcal Isolates from Infants and Pregnant Women in Portugal Revealing Two Lineages with Enhanced Invasiveness. J Clin Microbiol. **45**:3224-3229.
89. **Imperi, M., G. Gherardi, A. Berardi, L. Baldassarri, M. Pataracchia, G. Dicuonzo, G. Orefici, and R. Creti.** 2011. Invasive neonatal GBS infections from an area-based surveillance study in Italy. Clin Microbiol Infect. **17**:1834-1839.
90. **Bellais, S., A. Six, A. Fouet, M. Longo, N. Dmytruk, P. Glaser, P. Trieu-Cuot, and C. Poyart.** 2012. Capsular Switching in Group B *Streptococcus* CC17 Hypervirulent Clone: A Future Challenge for Polysaccharide Vaccine Development. J Infect Dis. **206**:1745-1752.

91. **Amin, A., Y. M. Abdulrazzaq, and S. Uduman.** 2002. Group B Streptococcal Serotype Distribution of Isolates from Colonized Pregnant Women at the Time of Delivery in United Arab Emirates. J Infect. **45**:42-46.
92. **Diedrick, M. J., A. E. Flores, S. L. Hillier, R. Creti, and Patricia Ferrieri.** 2010. Clonal Analysis of Colonizing Group B *Streptococcus*, Serotype IV, an Emerging Pathogen in the United States. J Clin Microbiol. **48**:3100-3104.
93. **Ferrieri, P., R. Lynfield, R. Creti, and A. Flores**. 2013. Serotype IV and Invasive Group B *Streptococcus* Disease in Neonates. Emerg Infect Dis. **19**:551-558.
94. **Palmeiro, J. K., L. M. Dalla-Costa, S. E. L. Fracalanzza, A. C. N. Botelho, K. S. Nogueira, M. C. Scheffer, R. S. L. A. Torres, N. S. Carvalho, L. L. Cogo, and H. M. F. Madeira.** 2010. Phenotypic and Genotypic Characterization of Group B Streptococcal Isolates in Southern Brazil. J Clin Microbiol. **48**:4397-4403.
95. **Skoff, T. H., M. M. Farley, S. Petit, A. S. Craig, W. Schaffner, K. Gershman, L. H. Harrison, R. Lynfield, J. Mohle-Boetani, S. Zansky, B. A. Albanese, K. Stefonek, E. R. Zell, D. Jackson, T. Thompson, and S. J. Schrag.** 2009. Increasing Burden of Invasive Group B Streptococcal Disease in Nonpregnant Adults, 1990-2007. Clin Infect Dis. **49**:85-92.
96. **Tien, N., C. Ho, H. Lin, M. Shih, M. Ho, H. Lin, H. Lin, C, Chang, and J. Lu.** 2011. Multilocus sequence typing of invasive group B *Streptococcus* in central area of Taiwan. J Microbiol, Immunol and Infect. **44**:430-434.
97. **Teatero, S., A. McGeer, A. Li, J. Gomes, C. Seah, W. Demczuk, I. Martin, J. Wasserscheid, K. Dewar, R. G. Melano, N. Fittipaldi.** 2015. Population Structure and Antimicrobial Resistance of Invasive Serotype IV Group B *Streptococcus*, Toronto, Ontario, Canada. Emerg Infect Dis. **21**(4):585-591.
98. **Foxman, B., L. Zhang, J. S. Koopman, S. D. Manning, and C. F. Marrs.** 2005. Choosing an appropriate bacterial typing technique for epidemiologic studies. Epidemiologic Perspectives & Innovations. **2**:10.
99. **Jones, N., J. F. Bohnsack, S. Takahashi, K. A. Oliver, M. S. Chan, F. Kunst, P. Glaser, C. Rusniok, D. W. Crook, R. M. Harding, N. Bisharat, and B. G. Spratt.** 2003. Multilocus sequence typing system for group B *Streptococcus*. J Clin Microbiol. **41**:2530-2536.
100. **Melles, D. C., W. B. van Leeuwen, S. V. Snijders, D. Horst-Kreft, J. K. Peeters, H. A. Verbrugh, and A. van Belkum.** 2007. Comparison of multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and amplified fragment length polymorphism (AFLP) for genetic typing of *Staphylococcus aureus*. Journal of Microbiological Methods. **69**:371-375.
101. **Florindo, C., J. P. Gomes, M. G. Rato, L. Bernardino, B. Spellerberg, I. Santos-Sanches, and M. J. Borrego.** 2011. Molecular epidemiology of group B streptococcal meningitis in children beyond the neonatal period from Angola. J Med Microbiol. **60**:1276-1280.
102. **Jones, N., K. A. Oliver, J. Barry, R. M. Harding, N. Bisharat, B. G. Spratt, T. Peto, and D. W. Crook.** 2006. Enhanced Invasiveness of Bovine-Derived Neonatal Sequence Type 17 Group B *Streptococcus* Is Independent of Capsular Serotype. Clin Infect Dis. **42**:915-924.
103. **Lamy, M., S. Dramsi, A. Billoet, H. Réglier-Poupet, A. Tazi , J. Raymond, F. Guérin, E. Couvé, F. Kunst, P. Glaser , P. Trieu-Cuot, and C. Poyart.** 2006. Rapid detection of the ''highly virulent'' group B *Streptococcus* ST-17 clone. Microb Infect. **8**:1714-1722.
104. **Lopez-Sanchez, M., E. Sauvage, V. Da Cunha, D. Clermont, E. R. Hariniaina, B. Gonzalez-Zorn, C. Poyart, I. Rosinski-Chupin, and P. Glaser.** 2012. The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. Mol Microbiol. **85**:1057-1071.
105. **Lier, C., E. Baticle, P. Horvath, E. Haguenoer, A. Valentin, P. Glaser, L. Mereghetti, and P. Lanotte.** 2015. Analysis of the type II-A CRISPR-Cas system of *Streptococcus agalactiae* reveals distinctive features according to genetic lineages. Frontiers in Genetics. **6**:214.
106. **Armistead, B., E. Oler, K. A. Waldorf, and L. Rajagopal.** 2019. The Double Life of Group B *Streptococcus*: Asymptomatic Colonizer and Potent Pathogen. J Mol Biol. **431**:2914-31.
107. **Gori, A., O. B. Harrison, E. Mlia, Y. Nishihara, J. M. Chan, J. Msefula, M. Mallewa, Q. Dube, T. D. Swarthout, A. H. Nobbs, M. C. J. Maiden, N. French, R. S. Heydermana.**

2020. Pan-GWAS of *Streptococcus agalactiae* Highlights LineageSpecific Genes Associated with Virulence and Niche Adaptation. mBio. **11**(3): e00728-20.
108. **Pietrocola, G., C. R. Arciola, S. Rindi, L. Montanaro, and P. Speziale.** 2018. *Streptococcus agalactiae* Non-Pilus, Cell Wall-Anchored Proteins: Involvement in Colonization and Pathogenesis and Potential as Vaccine Candidates. Front Immunol. **9**(602).
109. **Thomas, L., and L. Cook.** 2020. Two-Component Signal Transduction Systems in the Human Pathogen *Streptococcus agalactiae*. Infection and Immunity. **88**(7):e00931-19.
110. **Patras, K.A., J. Derieux, M. M. Al-Bassam N. Adiletta, A. Vrbanac, J. D. Lapek, K. Zengler, D. J. Gonzalez, and V. Nizet.** 2018. Group B Streptococcus biofilm regulatory protein A contributes to bacterial physiology and innate immune resistance. J Infect Dis. **218**(10):1641-1652.
111. **Tettelin, H., V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. Peterson, M. R. Wessels, I. T. Paulsen, K. E. Nelson, I. Margarit, T. D. Read, L. C. Madoff, A. M. Wolf, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, R. T. DeBoy, A. 64 S. Durkin, J. F. Kolonay, R. Madupu, M. R. Lewis, D. Radune, N. B. Fedorova, D. Scanlan, H. Khouri, S. Mulligan, H. A. Carty, R. T. Cline, S. E. Van Aken, J. Gill, M. Scarselli, M. Mora, E. T. Iacobini, C. Brettoni, G. Galli, M. Mariani, F. Vegni, D. Maione, D. Rinaudo, R. Rappuoli, J. L. Telford, D. L. Kasper, G. Grandi, and C. M. Fraser.** 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. PNAS. **99**:12391-12396.
112. **Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. DeBoy, T. M. Davidsen, M. Mora, M. Scarselli, I. M. Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. B. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli, and C. M. Fraser.** 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". PNAS. **102**:13950-13955.
113. **Richards, V.P., P. Lang, P.D.P. Bitar, T. Leféburea, Y. H. Schukkenb, R. N. Zadoksb, and M. J. Stanhope.** 2011. Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*. Infect Genet Evol. **11**:1263-1275.
114. **Wang, S., X. Dong, Y. Zhu, C. Wang, G. Sun, T. Luo, W. Tian, H. Zheng, and Q. Gao.** 2013. Revealing of *Mycobacterium marinum* Transcriptome by RNA-seq. PLoS One. **8**(9):e75828.
115. **Lowe, R., N. Shirley, M. Bleackley, S. Dolan, and T. Shafee.** 2017. Transcriptomics technologies. PLoS Comput Biol. **13**(5):e1005457.
116. **Richards, V. P., S. C. Choi, P. D. P. Bitar, A. A. Gurjar, and M. J. Stanhope.** 2013. Transcriptomic and genomic evidence for *Streptococcus agalactiae* adaptation to the bovine environment. BMC Genomics. **14**:920.
117. **Rosinski-Chupin, I., E. Sauvage, O. Sismeiro, A. Villain, V. Da Cunha, M. Caliot, M. Dillies, P. Trieu-Cuot, P. Bouloc, M. Lartigue, and P. Glaser.** 2015. Single nucleotide resolution RNA-seq uncovers new regulatory mechanisms in the opportunistic pathogen *Streptococcus agalactiae*. BMC Genomics. **16**:419.
118. **Rosinski-Chupin, I., E. Sauvage, A. Fouet, C. Poyart, and P. Glaser.** 2019. Conserved and specific features of *Streptococcus pyogenes* and *Streptococcus agalactiae* transcriptional landscapes. BMC Genomics. **20**:236.
119. **Spencer, B. L., L. Deng, K. A. Patras, Z. M. Burcham, G. F. Sanches, P. E. Nagao, and K. S. Doran.** 2019. Cas9 Contributes to Group B Streptococcal Colonization and Disease. Front Microbiol. **10**:1930.
120. **Hooven, T. A., A. J. Catomeris, M. Bonakdar, L. J. Tallon, I. Santana-Cruz, S. Ott, S. C. Daugherty, H. Tettelin, and A. J. Ratner.** 2018. The *Streptococcus agalactiae* Stringent Response Enhances Virulence and Persistence in Human Blood. Infection and Immunity. **86**(1):e00612-17.

121. **Cook, L. C. C., H. Hu, M. Maienschein-Cline, and M. J. Federle.** 2018. A Vaginal Tract Signal Detected by the Group B Streptococcus SaeRS System Elicits Transcriptomic Changes and Enhances Murine Colonization. Infection and Immunity. **86**(4):e00762-17.
122. **Lin, F. C., A. Whiting, E. Adderson, S. Takahashi, D. M. Dunn, R. Weiss, P. H. Azimi, J. B. Philips III, L. E. Weisman, J. Regan, P. Clark, G. G. Rhoads, C. E. Frasch, J. Troendle, P. Moyer, and J. F. Bohnsack.** 2006. Phylogenetic Lineages of Invasive and Colonizing Strains of Serotype III Group B Streptococci from Neonates: a Multicenter Prospective Study. J Clin Microbiol. **44**:1257-1261.
123. **Almeida, A., I. Rosinski-Chupin, C. Plainvert, P. Douarre, M. J. Borrego, C. Poyart, and P. Glaser.** 2017. Parallel Evolution of Group B *Streptococcus* Hypervirulent Clonal Complex 17 Unveils New Pathoadaptive Mutations. mSystems. **2**(5):e00074-17.
124. **Manning, S. D., A. C. Springman, E. Lehotzky, M. A. Lewis, T. S. Whittam, and H. D. Davies.** 2009 Multilocus Sequence Types Associated with Neonatal Group B Streptococcal Sepsis and Meningitis in Canada. J Clin Microbiol. **47**(4):1143-1148.
125. **Seo, H.S., G. Minasov, R. Seepersaud, K. S. Doran, I. Dubrovska, L. Shuvalova, W. F. Anderson, T. M. Iverson, and P. M. Sullam.** 2013. Characterization of Fibrinogen Binding by Glycoproteins Srr1 and Srr2 of *Streptococcus agalactiae*. J Biol Chem. **288**(50):35982-35996.
126. **D'Urzo, N., M. Martinelli, A. Pezzicoli, V. De Cesare, V. Pinto, I. Margarit, J. L. Telford, D. Maione, and Members of the DEVANI Study Group.** 2014. Acidic pH strongly enhances in vitro biofilm formation by a subset of hypervirulent ST-17 *Streptococcus agalactiae* strains. Appl Environ Microbiol. **80**(7):2176-2185.
127. **Silvestre, I., M. J. Borrego, and L. Jordão.** 2020. Biofilm formation by ST17 and ST19 strains of *Streptococcus agalactiae*. Research in Microbiology. **171**(8):311-318.
128. **Florindo, C., C. A. Barroco, I. Silvestre, V. Damião, J.P. Gomes, B. Spellerberg, I. Santos-Sanches, and M. J. Borrego.** 2018. Capsular Type, Sequence Type and Microbial Resistance Factors Impact on DNase Activity of *Streptococcus agalactiae* Strains from Human and Bovine Origin. Eur J Microbiol Immunol. **8**(4):149-154.
129. **Rajagopal, L.** 2009. Understanding the regulation of Group B Streptococcal virulence factors. Future Microbiol. **4**(2): 201-221.
130. **Xia, F.D., A. Mallet, E. Caliot, C. Gao, P. Trieu-Cuot, and S. Dramsi.** 2015. Capsular polysaccharide of Group B *Streptococcus* mediates biofilm formation in the presence of human plasma. Microbes Infect. **17**(1):71-76.
131. **Teatero, S., P. Ferrieri, I. Martin, W. Demczuk, A. McGeer, N. Fittipaldi, and for the Toronto Invasive Bacterial Diseases Network.** 2017. Serotype Distribution, Population Structure, and Antimicrobial Resistance of Group B *Streptococcus* Strains Recovered from Colonized Pregnant Women. J Clin Microbiol. **55**(2):412-422.
132. **Teatero, S., P. Ferrieri, and N. Fittipaldi.** 2016. Serotype IV Sequence Type 468 Group B Streptococcus Neonatal Invasive Disease, Minnesota, USA. Emerg Infect Dis. 22(11):1937-1940.
133. **Whidbey, C., M. I. Harrell, K. Burnside, L. Ngo, A. K. Becraft, L. M. Iyer, L. Aravind, J. Hitti, K. M. A. Waldorf, and L. Rajagopal.** 2013. A hemolytic pigment of Group B *Streptococcus* allows bacterial penetration of human placenta. JEM. **210**(6):1265-81.
134. **Vornhagen, J., K. M. A. Waldorf, and L. Rajagopal.** 2017. Perinatal Group B Streptococcal Infections: Virulence Factors, Immunity and Prevention Strategies. Trends Microbiol. **25**(11):919-931.
135. **Rosa-Fraile, M., S. Dramsi, and B. Spellerberg.** 2014. Group B streptococcal haemolysin and pigment, a tale of twins. FEMS Microbiol Rev. **38**:932-946.
136. **Boldenow, E., C. Gendrin, L. Ngo, C. Bierle, J. Vornhagen, M. Coleman, S. Merillat, B. Armistead, C. Whidbey, V. Alishetti, V. Santana-Ufret, J. Ogle, M. Gough, S. Srinouanprachanh, J. W. MacDonald, T. K. Bammler, A. Bansal, H. D. Liggitt, L. Rajagopal, and K. M. A. Waldorf.** 2016. Group B *Streptococcus* circumvents neutrophils and neutrophil extracellular traps during amniotic cavity invasion and preterm labor. Sci Immunol. **1**(4).

137. **Berends, E. T. M., A. R. Horswill, N. M. Haste, M. Monestier, V. Nizet, and M. Kockritz-Blickwede.** 2010. Nuclease Expression by S*taphylococcus aureus* Facilitates Escape from Neutrophil Extracellular Traps. Journal of Innate Immunity. **2**:576-586.
138. **Remmington A., and C. E. Turner.** 2018. The DNases of pathogenic Lancefield streptococci. Microbiol **164:**242-250.
139. **Maisey, H.C., K. S. Doran, and V. Nizet.** 2008. Recent advances in understanding the molecular basis of group B *Streptococcus* virulence. Expert Rev Mol Med. **10**:e27.
140. **Gendrin, C., J. Vornhagen, B. Armistead, P. Singh, C. Whidbey, S. Merillat, D. Knupp, R. Parker, L. M. Rogers, P. Quach, L. M. Iyer, L. Aravind, S. D. Manning, D. M. Aronoff, and L. Rajagopal.** 2018. A Nonhemolytic Group B Streptococcus Strain Exhibits Hypervirulence. J Infect Dis. **217**:983-987.
141. **Tazi, A., S. Bellais, I. Tardieux, S. Dramsi, P. Trieu-Cuot, and C. Poyart.** 2012. Group B *Streptococcus* surface proteins as major determinants for meningeal tropism. Current Opinion in Microbiology. **15**:44-49.
142. **Vilar, R., R. J. Fish, A. Casini, and M. Neerman-Arbez.** 2020. Fibrin(ogen) in human disease: both friend and foe. Haematologica. **105**(2):284-296.
143. **Buscetta, M., S. Papasergi, A. Firon, G. Pietrocola, C. Biondo, G. Mancuso, A. Midiri, L. Romeo, G. Teti, P. Speziale, P. Trieu-Cuot, and C. Beninati.** 2014. FbsC, a Novel Fibrinogen-binding Protein, Promotes *Streptococcus agalactiae*-Host Cell Interactions. J Biol Chem. **289**(30):21003-21015.
144. **Al Safadi, R., L. Mereghetti, M. Salloum, M. Lartigue, I. Virlogeux-Payant, R. Quentin, and A. Rosenau.** 2011. Two-Component System RgfA/C Activates the *fbsB* Gene Encoding Major Fibrinogen-Binding Protein in Highly Virulent CC17 Clone Group B Streptococcus. PLoS One. **6**(2):e14658.
145. **Burcham, L. R., B. L. Spencer, L. R. Keeler, D. L. Runft, K. A. Patras, M. N. Neely, and K. S. Doran.** 2019. Determinants of Group B streptococcal virulence potential amongst vaginal clinical isolates from pregnant women. PLoS One. **14**(12):e0226699.
146. **Al-Obaidi, M. M. J., and M. N. M. Desa.** 2018. Mechanisms of Blood Brain Barrier Disruption by Different Types of Bacteria, and Bacterial–Host Interactions Facilitate the Bacterial Pathogen Invading the Brain. Cell Mol Neurobiol. **38**:1349-1368.
147. **Tazi, A., O. Disson, S. Bellais, A. Bouaboud, N. Dmytruk, S. Dramsi, M. Mistou, H. Khun, C. Mechler, I. Tardieux, P. Trieu-Cuot, M. Lecuit, and C. Poyart.** 2010. The surface protein HvgA mediates group B S*treptococcus* hypervirulence and meningeal tropism in neonates. JEM. **207**:2313-2322.
148. **Broker, G., and B. Spellerberg.** 2004. Surface proteins of *Streptococcus agalactiae* and horizontal gene transfer. Int J Med Microbiol. **294**:169-175.
149. **Creti, R., F. Fabretti, G. Orefici, and C. Hunolstein.** 2004. Multiplex PCR Assay for direct Identification of Group B *Streptococcal* Alpha-Protein-Like Protein Genes. J Clin Microbiol. **42**:1326-1329.
150. **Maeland, J.A., J. E. Afset, R. V. Lyng, and A. Radtkea.** 2015. Survey of Immunological Features of the Alpha-Like Proteins of *Streptococcus agalactiae*. Clinical and Vaccine Immunology. **22**(2):153-159.
151. **Lachenauer, C. S., R. Creti, J. L. Michel, and L. C. Madoff.** 2000. Mosaicism in the alpha-like protein genes of group B streptococci. PNAS. **97**(17):9630-9635.
152. **Gabrielsen, C., J. A. Maeland, R. V. Lyng, A. Radtke, and J. E. Afset.** 2017. Molecular characteristics of *Streptococcus agalactiae* strains deficient in alpha-like protein encoding genes. J Med Microbiol. **66**:26–33.
153. **Brochet, M., E. Couvé, M. Zouine, T. Vallaeys, C. Rusniok, M. Lamy, C. Buchrieser, P. Trieu-Cuot, F. Kunst, C. Poyart, and P. Glaser.** 2006. Genomic diversity and evolution within the species *Streptococcus agalactiae*. Microb Infect. **8**:1227-1243.
154. **Paoletti, L. C., and D. L. Kasper.** 2019. Surface Structures of Group B *Streptococcus* Important in Human Immunity. Microbiol Spectrum. **7**(2).
155. **Jiang, S., S. E. Park, P. Yadav, L.C. Paoletti, and M. R. Wessels.** 2012. Regulation and Function of Pilus Island 1 in Group B *Streptococcus*. J Bacteriol. 194(10):2479-2490.

156. **Konto-Ghiorghi, Y., E. Mairey E., A. Mallet, D. Guillaume, E. Caliot, P. Trieu-Cuot, and S. Dramsi.** 2009. Dual Role for Pilus in Adherence to Epithelial Cells and Biofilm Formation in *Streptococcus agalactiae*. PLoS Pathog. 5(5):e1000422.
157. **Rinaudo C.D., Rosini R., Galeotti C.L., F. Berti, F. Necchi, V. Reguzzi, C. Ghezzo, J. L. Telford, G. Grandi, and D. Maione.** 2010. Specific involvement of pilus type 2a in biofilm formation in group B *Streptococcus*. PLoS One. **5**(2):e9216.
158. **Lazzarin, M., R. Mu, M. Fabbrini, C. Ghezzo, C. D. Rinaudo, K. S. Doran, and I. Margarit.** 2017. Contribution of pilus type 2b to invasive disease caused by a *Streptococcus agalactiae* ST-17 strain. BMC Microbiol. **17**:148.
159. **Lu, B., J. Wu, X. Chen, C. Gao, J. Yang, Y. Li, J. Wang, J. Zeng, Y. Fang, D. Wang, Y. Cui, and L. Wang.** 2018. Microbiological and clinical characteristics of Group B *Streptococcus* isolates causing materno-neonatal infections: high prevalence of CC17/PI-1 and PI-2b sublineage in neonatal infections. J Med Microbiol. **67**:1551-1559.
160. **Lu, B., D. Wang, H. Zhou, F. Zhu, D. Li, S. Zhang, Y. Shi, Y. Cui, L. Huang, and H. Wu.** 2015. Distribution of pilus islands and alpha-like protein genes of group B *Streptococcus* colonized in pregnant women in Beijing, China. Eur J Clin Microbiol Infect Dis. **34**:1173-1179.
161. **Campisi, E., R. Rosini, W. Ji, S. Guidotti, M. Rojas-López, G. Geng, Q. Deng, H. Zhong, W. Wang, H. Liu, C. Nan, I. Margarit, and C. D. Rinaudo.** 2016. Genomic Analysis Reveals Multi-Drug Resistance Clusters in Group B Streptococcus CC17 Hypervirulent Isolates Causing Neonatal Invasive Disease in Southern Mainland China. Front Microbiol. **7**:1265.
162. **Ferrieri, P., E. D. Gray, and L. W. Wannamaker.** 1980. Biochemical and immunological characterization of the extracellular nucleases of Group B streptococci. J Experiment Med. **151**:56-68.
163. **Derré-Bobillot, A., N. G. Cortes-Perez, Y. Yamamoto, P. Kharrat, E. Couvé, V. Da Cunha, P. Decker, M. Boissier, F. Escartin, B. Cesselin, P. Langella, L. G. Bermúdez-Humarán, and P. Gaudu.** 2013. Nuclease A (Gbs0661), an extracellular nuclease of *Streptococcus agalactiae*, attacks the neutrophil extracellular traps and is needed for full virulence. Mol Microbiol. **89**(3):518–531.
164. **Brinkmann, V., U. Reichard, C. Goosmann, B. Fauler, Y. Uhlemann, D. S. Weiss, Y. Weinrauch, and A. Zychlinsky.** 2004. Neutrophil Extracellular Traps Kill Bacteria. Science. **303**:1532-1535.
165. **Brinkmann, V., and A. Zychlinsky.** 2012. Neutrophil extracellular traps: Is immunity the second function of chromatin? JCB. **198**(5):773-783.
166. **Aziz, R. K., R. A. Edwards, W. W. Taylor, D. E. Low, A. McGeer, and M. Kotb.** 2005. Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated M1T1 clone of *Streptococcus pyogenes*. J Bacteriol. **187**:3311-3318.
167. **Maamary, P.G., N. L. B. Zakour, J. N. Cole, A. Hollands, R. K. Aziz, T. C. Barnett, A. J. Cork, A. Henningham, M. Sanderson-Smith, J. D. McArthur, C. Venturini, C. M. Gillen, J. K. Kirk, D. R. Johnson, W. L. Taylor, E. L. Kaplan, M. Kotb, V. Nizet, S. A. Beatson, and M. J. Walker.** 2012. Tracing the evolutionary history of the pandemic group A streptococcal M1T1 clone. FASEB J. **26**:4675-4684.
168. **Rosini, R., and I. Margarit.** 2015. Biofilm formation by *Streptococcus agalactiae*: influence of environmental conditions and implicated virulence factors. Front Cell Infect Microbiol. **5**:6.
169. **Nie, S., X. Lu, Y. W. Hu, L. Zheng, and Q. Wang.** 2018. Influence of environmental and genotypic factors on biofilm formation by clinical isolates of group B *streptococci*. Microb Pathog. **121**:45-50.
170. **Parker, R. E., C. Laut, J. A. Gaddy, R. N. Zadoks, H. D. Davies, and S. D. Manning.** 2016. Association between genotypic diversity and biofilm production in group B *Streptococcus*. BMC Microbiol. **16**:86.
171. **Borges, S., J. Silva, and P. Teixeira.** 2012. Survival and biofilm formation by Group B *Streptococci* in simulated vaginal fluid at different pHs. Antonie van Leeuwenhoek. **101**:677-682.

172. **Ho, Y.R., C. M. Li, C. H. Yu, Y. J. Lin, C. M. Wu, I. C. Harn, M. J. Tang, Y. T. Chen, F. C. Shen, C. Y. Lu, T. C. Tsai, and J. J. Wu.** 2013. The enhancement of biofilm formation in Group B streptococcal isolates at vaginal pH. Med Microbiol Immunol. **202**(2):105-115.
173. **Kaur, H., P. Kumar, P. Ray, J. Kaur, and A. Chakraborti.** 2009. Biofilm formation in clinical isolates of group B *Streptococci* from north India. Microb Pathog. **46**(6):321-327.
174. **Yang, Q., A. J. Porter, M. Zhang, D. J. Harrington, G. W. Black, and I. C. Sutcliffe.** 2012. The impact of pH and nutrient stress on the growth and survival of *Streptococcus agalactiae*. Antonie van Leeuwenhoek. **102**(2):277-87.
175. **Spellerberg, B.** 2000. Pathogenesis of neonatal *Streptococcus agalactiae* infections. Microb Infect. **2**:1733-1742.
176. **Storisteanu, D., J. Pocock, A. Cowburn, J. Juss, A. Nadesalingam, V. Nizet, and E. Chilvers.** 2017. Evasion of Neutrophil Extracellular Traps by Respiratory Pathogens. Am J Respir Cell Mol Biol. **56**(4):423–431.
177. **Liu, J., L. Sun, W. Liu, L. Guo, Z. Liu, X. Wei, and J. Ling.** 2017. A Nuclease from *Streptococcus mutans* Facilitates Biofilm Dispersal and Escape from Killing by Neutrophil Extracellular Traps. Front Cell Infect Microbiol. **7**:97.
178. **Kao, Y., M. Tsai, M. Lai, S. Chu, H. Huang, M. Chiang, R. Fu, J. Lu, and J. Hsu.** 2019. Emerging serotype III sequence type 17 group B streptococcus invasive infection in infants: the clinical characteristics and impacts on outcomes. BMC Infect Dis. **19**:538.
179. **Sumby, P., K. D. Barbian, D. J. Gardner, A. R. Whitney, D. M. Welty, R. D. Long, J. R. Bailey, M. J. Parnell, N. P. Hoe, G. G. Adams, F. R. DeLeo, and J. M. Musser.** 2005. Extracellular deoxyribonuclease made by group A *Streptococcus* assists pathogenesis by enhancing evasion of the innate immune response. PNAS. **102**(5):1679-1684.
180. **Florindo, C., R. Ferreira, V. Borges, B. Spellerberg, J.P. Gomes, and M.J. Borrego.** 2012. Selection of reference genes for real-time expression studies in *Streptococcus agalactiae*. Journal of Microbiological Methods. **90**(3):220-227.
181. **Pelerito, A., A. Nunes, M. S. Núncio, and J. P. Gomes.** 2020. Genome-scale approach to study the genetic relatedness among Brucella melitensis strains. PLoS One. **15**(3):e0229863.
182. **Llarena, A., B. F. Ribeiro-Gonçalves, D. N. Silva, J. Halkilahti, M. P.Machado, M.S. Da Silva, A. Jaakkonen, J. Isidro, C. Hämäläinen, J. Joenperä, V. Borges, L. Viera, J. P. Gomes, C. Correia, J. Lunden, R. Laukkanen-Ninios, M. Fredriksson-Ahomaa, J. Bikandi, R. San Millan, I. Martinez-Ballesteros, L. Laorden, M. Mäesaar, L. Grantina-Ievina, F. Hilbert, J. Garaizar, M. Oleastro, M. Nevas, S. Salmenlinna, M. Hakkinen, J. A. Carriço, and M. Rossi.** 2018. INNUENDO: A cross-sectoral platform for the integration of genomics in the surveillance of food-borne pathogens. EFSA Support Publ. **15**(11):1498E.
183. **Bolger, A. M., M. Lohse, and B. Usadel.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. **30**(15):2114-2120.
184. **Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. **19**(5):455-477.
185. **Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, and A. M. Earl.** 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One. **9**(11):e112963.
186. **Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy.** 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. **27**:722-736.
187. **Li, H.** 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. **34**:3094-3100.
188. **Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R.**

Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. **9**:75.
189. **Seemann, T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. **30**:2068-2069.
190. **Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov.** 2011. Integrative Genomics Viewer. Nature Biotechnology. **29**:24-26.
191. **Kumar, S., G. Stecher, and K. Tamura.** 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution. **33**:1870-1874.
192. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. **4**:406-425.
193. **Tamura, K., M. Nei, and S. Kumar.** 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA). **101**:11030-11035.
194. **Felsenstein, J.** 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. **39**:783-791.
195. **Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. **186**:1518-1530.
196. **Zhang, Q., and Y. Ye.** 2017. Not all predicted CRISPR-Cas systems are equal: Isolated cas genes and classes of CRISPR like elements. BMC Bioinformatics. **18**:92.
197. **Chen, L., J. Yang, J. Yu, Z. Yao, L. Sun, Y. Shen, and Q. Jin.** 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acid Res. **33**(Database issue):D325-328.
198. **Arndt, D., J. R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, and D. S. Wishart.** 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. **44**(W1):W16-21.
199. **Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. G. Holden, M. Fookes, D. Falush, J. A. Keane, and J. Parkhill.** 2015. Bioinformatics. **31**(22):3691-3693.
200. **Moon, A. F., P. Gaudub, and L. C. Pedersen.** 2014. Structural characterization of the virulence factor nuclease A from *Streptococcus agalactiae*. Acta Cryst. **D70:**2937–2949.
201. **Erill, I., S. Campoy, and J. Barbé.** 2007. Aeons of distress: an evolutionary perspective on the bacterial SOS response. FEMS Microbiol Rev. **31**:637-656.
202. **Boutry, C., B. Delplace, A. Clippe, L. Fontaine, and P. Hols.** 2013. SOS Response Activation and Competence Development Are Antagonistic Mechanisms in Streptococcus thermophilus. J Bacteriol. **195**(4):696-707.
203. **Mijakovic, I., D. Petranovic, B. Macek, T. Cepo, M. Mann, J. Davies, P. R. Jensen, and D. Vujaklija.** 2006. Bacterial single-stranded DNA-binding proteins are phosphorylated on tyrosine. Nucleic Acids Research. **34**(5):1588-1596.
204. **Spenkelink, L. M., J. S. Lewis, S. Jergic, Z. Xu, A. Robinson, N. E. Dixon, and A. M. van Oijen.** 2019. Recycling of single-stranded DNA-binding protein by the bacterial replisome. Nucleic Acids Research. **47**(8):4111-4123.
205. **Guérillot, R., V. Da Cunha, E. Sauvage, C. Bouchier, and P. Glaser.** 2013. Modular Evolution of TnGBSs, a New Family of Integrative and Conjugative Elements Associating Insertion Sequence Transposition, Plasmid Replication, and Conjugation for Their Spreading. J Bacteriol. **195**(9):1979-1990.
206. **Sutcliffe, I. C., G. W. Black, and D. J. Harrington.** 2008. Bioinformatic insights into the biosynthesis of the Group B carbohydrate in *Streptococcus agalactiae*. Microbiol. **154**:1354-1363.
207. **Roberts, R. J., M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. K. Degtyarev, D. T. F. Dryden, K. Dybvig, K. Firman, E. S. Gromova, R. I. Gumport, S. E. Halford, S. Hattman, J. Heitman, D. P. Hornby, A. Janulaitis, A. Jeltsch, J. Josephsen, A. Kiss, T. R. Klaenhammer, I. Kobayashi, H. Kong, D. H. Kruger, S. Lacks, M. G. Marinus, M. Miyahara, R. D. Morgan, N. E. Murray, V. Nagaraja, A. Piekarowicz, A. Pingoud, E. Raleigh, D. N. Rao, N. Reich, V. E. Repin, E. U. Selker, P. Shaw, D. C. Stein, B.**

L. Stoddard, W. Szybalski, T. A. Trautner, J. L. Van Etten, J. M. B. Vitor, G. G. Wilson, and S. Xu. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Research. **31**(7):1805-1812.

208. **Kumar, S., X. Cheng, S. Klimasauskas, M. Sha, J. Posfai, R. J. Roberts, and G. G. Wilson.** 1994. The DNA (cytosine-5) methyltransferases. Nucleic Acids Research. **22**(1):1-10.

209. **Paetzel, M., A. Karla, N. C. J. Strynadka, and R. E. Dalbey.** 2002. Signal peptidases. Chem Rev. **102**(12):4549-4580.

210. **Owji, H., N. Nezafat, M. Negahdaripour, A. Hajiebrahimi, and Y. Ghasemi.** 2018. A comprehensive review of signal peptides: Structure, roles, and applications. Eur J Cell Biol. **97**(6):422-441.

211. **Frees, D., A. Chastanet, S. Qazi, K. Sørensen, P. Hill, T. Msadek, and H. Ingmer.** 2004. Clp ATPases are required for stress tolerance, intracellular replication and biofilm formation in Staphylococcus aureus. Mol Microbiol. **54**(5):1445-1462.

212. **Nair, S., C. Poyart, J. Beretti, H. Veiga-Fernandes, P. Berche, and P. Trieu-Cuot.** 2003. Role of the *Streptococcus agalactiae* ClpP serine protease in heat-induced stress defence and growth arrest Shamila. Microbiol. **149**:407-417.

213. **Aguilar-Rodríguez, J., B. Sabater-Muñoz, R. Montagud-Martínez, V. Berlanga, D. Alvarez-Ponce, A. Wagner, and M. A. Fares.** 2016. The Molecular Chaperone DnaK Is a Source of Mutational Robustness. Genome Biol Evol. **8**(9):2979-2991.

214. **Chiappori, F., M. Fumian, L. Milanesi, and I. Merelli.** 2015. DnaK as Antibiotic Target: Hot Spot Residues Analysis for Differential Inhibition of the Bacterial Protein in Comparison with the Human HSP70. PLoS One. **10**(4):e0124563.

215. **Tomoyasu, T., A. Tabata, H. Imaki, K. Tsuruno, A. Miyazaki, K. Sonomoto, R. A. Whiley, and H. Nagamune.** 2012. Role of Streptococcus intermedius DnaK chaperone system in stress tolerance and pathogenicity. Cell Stress and Chaperones. **17**:41-55.

216. **Grove, T. L., J. S. Benner, M. I. Radle, J. H. Ahlum, B. J. Landgraf, C. Krebs, and S. J. Booker.** 2011. A Radically Different Mechanism for S-Adenosylmethionine–Dependent Methyltransferases. Science. **332**(6029):604-607.

217. **Kozbial, P. Z., and A. R. Mushegian.** 2005. Natural history of S-adenosylmethionine-binding proteins. BMC Structural Biology. **5**:19.

218. **Struck, A., M. L. Thompson, L. S. Wong, and J. Micklefield.** 2012. S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. ChemBioChem. **13**:2642-2655.

219. **Bonventre, J. A., R. A. Zielke, K. V. Korotkov, and A. E. Sikora.** 2016. Targeting an Essential GTPase Obg for the Development of Broad-Spectrum Antibiotics. PLoS One. **11**(2):e0148222.

220. **Grossman, L., and A. T. Yeung.** 1990. The UvrABC endonuclease system of Escherichia coli — A view from Baltimore. Mutation Research/DNA Repair. **236**(2-3):213-221.

221. **Ramos, J. L., M. Martínez-Bueno, A. J. Molina-Henares, W. Terán, K. Watanabe, X. Zhang, María T. Gallegos, R. Brennan, and R. Tobes.** 2005. The TetR Family of Transcriptional Repressors. Microbiol Mol Biol Rev. **69**(2):326-356.

222. **Cuthbertson, L., and J. R. Nodwell.** 2013. The TetR Family of Regulators. Microbiol Mol Biol Rev. **77**(3):440-475.

223. **Deng, W., L. Chunmei, and J. Li.** 2013. The underling mechanism of bacterial TetR/AcrR family transcriptional repressors. Cellular Signalling. **25**(7): 1608-1613.

224. **Rantanen, M. K., L. Lehtio, L. Rajagopal, C. E. Rubens, and A. Goldman.** 2007. Structure of *Streptococcus agalactiae* serine/threonine phosphatase. FEBS Journal. **274**:3128-3137.

225. **Shimizu, K.** 2013. Metabolic Regulation of a Bacterial Cell System with Emphasis on Escherichia coli Metabolism. ISRN Biochemistry. **2013**:645983.

226. **Willenborg, J., and R. Goethe.** 2016. Metabolic traits of pathogenic streptococci. FEBS Letters. **590**:3905–3919.

227. **Mee-Marquet, N., S. M. Diene, L. Barbera, L. Courtier-Martinez, L. Lafont, A. Ouachée, A. S. Valentin, S. Dos Santos, R. Quentin, and P. François.** 2018. Analysis of the

prophages carried by human infecting isolates provides new insight into the evolution of Group B Streptococcus species. Clin Microbiol Infect. **24**:514-521.
228. **Javan, R. R., E. Ramos-Sevillano, A. Akter, J. Brown, and A. B. Brueggemann.** 2019. Prophages and satellite prophages are widespread in Streptococcus and may play a role in pneumococcal pathogenesis. Nature Communications. **10**:4852.
229. **Zhang, W., C. Rong, C. Chen, and G. F. Gao.** 2012. Type-IVC Secretion System: A Novel Subclass of Type IV Secretion System (T4SS) Common Existing in Gram-Positive Genus Streptococcus. PLoS One. **7**(10):e46390.
230. **Navarre, W. W., and O. Schneewind.** 1994. Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in Gram-positive bacteria. Mol Microbiol. **14**(1):115-121.
231. **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**(1):33-40.
232. **Warne, B., C. P. Harkins, S. R. Harris, A. Vatsiou, N. Stanley-Wall, J. Parkhill, S. J. Peacock, T. Palmer, and M. T. G. Holden.** 2016. The Ess/Type VII secretion system of *Staphylococcus aureus* shows unexpected genetic diversity. BMC Genomics. **17**:222.
233. **Fléchard, M., and P. Gilot.** 2014. Physiological impact of transposable elements encoding DDE transposases in the environmental adaptation of *Streptococcus agalactiae*. Microbiol. **160**:1298-1315.
234. **Shabayek, S., and Spellerberg B.** 2017. Acid Stress Response Mechanisms of Group B *Streptococci*. Front Cell Infect Microbiol. **7**:395.
235. **Ferreira, R., V. Borges, M. J. Borrego, and J. P. Gomes.** 2017. Global survey of mRNA levels and decay rates of *Chlamydia trachomatis* trachoma and lymphogranuloma venereum biovars. Heliyon. **3**:e00364.
236. **Stark, R., M. Grzelak, and J. Hadfield.** 2019. RNA sequencing: the teenage years. Nature Reviews Genetics. **20**:631–656.
237. **Leimbach, A.** 2016. bac-Genomics-Scripts: Bovine, *E. coli* Mastitis Comparative Genomics Edition. Available online: https://zenodo.org/record/215824#.XMF5wmhL-Uk (accessed on 21 December 2016).
238. **Langmead, B., and S. L. Salzberg.** 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. **9**:357–359.
239. **Powell, D. R.** Degust: interactive RNA-seq analysis. No. DOI: 10.5281/zenodo.3258932.
240. **Bateman, A., and N. D. Rawlings.** 2003. The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. Trends in Biochemical Sciences. **28**(5):234-237.
241. **Reinscheid, D. J., K. Ehlert, G. S. Chhatwal, and B. J. Eikmanns.** 2003. Functional analysis of a PcsB-deficient mutant of group B streptococcus. FEMS Microbiol Lett. **221**:73-79.
242. **Reinscheid, D. J., B. Gottschalk, A. Schubert, B. J. Eikmanns, and G. S. Chhatwal.** 2001. Identification and Molecular Analysis of PcsB, a Protein Required for Cell Wall Separation of Group B Streptococcus. J Bacteriol. **183**(4):1175-1183.
243. **Zucchini, L., C. Mercy, P. S. Garcia, C. Cluzel, V. Gueguen-Chaignon, F. Galisson, C. Freton, S. Guiral, C. Brochier-Armanet, P. Gouet, and C. Grangeasse.** 2018. PASTA repeats of the protein kinase StkP interconnect cell constriction and separation of *Streptococcus pneumoniae*. Nature Microbiol. **3**:197-209.
244. **Roach, D. R., and D. M. Donovan.** 2015. Antimicrobial bacteriophage-derived proteins and therapeutic applications. Bacteriophage. **5**(3): e1062590-1.
245. **Mereghetti, L., I. Sitkiewicz, N. M. Green, and J. M. Musser.** 2008. Extensive Adaptive Changes Occur in the Transcriptome of *Streptococcus agalactiae* (Group B Streptococcus) in Response to Incubation with Human Blood. PLoS One. **3**(9):e3143.
246. **Mereghetti, L., I. Sitkiewicz, N. M. Green, and J. M. Musser.** 2008. Remodeling of the *Streptococcus agalactiae* Transcriptome in Response to Growth Temperature. PLoS One. **3**(7):e2785.

247. **Lanotte, P., M. Perivier, E. Haguenoer, L. Mereghetti, C. Burucoa, S. Claverol, and C. Atanassov.** 2013. Proteomic Biomarkers Associated with *Streptococcus agalactiae* Invasive Genogroups. PLoS One. **8**(1):e54393.
248. **Lamy, M., M. Zouine, J. Fert, M. Vergassola, E. Couve, E. Pellegrini, P. Glaser, F. Kunst, T. Msadek, P. Trieu-Cuot, and C. Poyart.** 2004. CovS/CovR of group B streptococcus: a two-component global regulatory system involved in virulence. Mol Microbiol. **54**(5):1250–1268.
249. **Ayres, C. A., N. Schormann, O. Senkovich, A. Fry, S. Banerjee, G. C. Ulette, and D. Chattopadhyay.** 2014. Structure of *Streptococcus agalactiae* glyceraldehyde-3-phosphate dehydrogenase holoenzyme reveals a novel surface. Acta Cryst. F70:1333–1339.
250. **Colón, D. F., C. W. Wanderley, M. Franchin, C. M. Silva, C. H. Hiroki, F. V. S. Castanheira, P. B. Donate, A. H. Lopes, L. C. Volpon, S. K. Kavaguti, V. F. Borges, C. A. Speck-Hernandez, F. Ramalho, A. P. Carlotti, F. Carmona, J. C. Alves-Filho, F. Y. Liew, and F. Q. Cunha.** 2019. Neutrophil extracellular traps (NETs) exacerbate severity of infant sepsis. Critical Care. **23**:113.
251. **Bandeira, M., V. Borges, J. P. Gomes, A. Duarte, and L. Jordão.** 2017. Insights on *Klebsiella pneumoniae* Biofilms Assembled on Different Surfaces Using Phenotypic and Genotypic Approaches. Microorganisms. **5**:16.
252. **Okshevsky M., and R. L. Meyer.** 2015. The role of extracellular DNA in the establishment, maintenance, and perpetuation of bacterial biofilms. Crit Rev Microbiol. **41**(3):341-352.
253. **Donelli, G.** 2014. Methods in Molecular Biology, vol 1147. Humana Press, New York.
254. **Nijland, R., M. J. Hall, and J. G. Burgess.** 2010. Dispersal of biofilms by secreted, matrix degrading, bacterial DNase. PLoS One. **5**(12):e15668.
255. **Seper, A., V. H. Fengler, S. Roier, H. Wolinski, S. D. Kohlwein, A. L. Bishop, A. Camilli, J. Reidl, and S. Schild.** 2011. Extracellular nucleases and extracellular DNA play important roles in *Vibrio cholerae* biofilm formation. Mol Microbiol. **82**(4):1015-1037.
256. **Florindo, C., S. Viegas, A. Paulino, E. Rodrigues, J. P. Gomes, and M. J. Borrego.** 2010. Molecular characterization and antimicrobial susceptibility profiles in *Streptococcus agalactiae* colonizing strains: association of erythromycin resistance with subtype III-I genetic clone family. Clin Microbiol Infect. **16**:1458-1463.
257. **Fluegge, K., J. Wons, B. Spellerberg B, S. Swoboda, A. Siedler, M. Hufnagel, and R. Berner.** 2011. Genetic differences between invasive and noninvasive neonatal group B streptococcal isolates. Pediatr Infect Dis J. **30**(11):1027-1031.
258. **Fluegge, K., S. Supper, A. Siedler, and R. Berner.** 2005. Serotype distribution of invasive group B streptococcal isolates in infants: results from a nationwide active laboratory surveillance study over 2 years in Germany. Clin Infect Dis. **40**(5):760-763.
259. **Fluegge, K., S. Supper, A. Siedler, and R. Berner.** 2004. Antibiotic susceptibility in neonatal invasive isolates of *Streptococcus agalactiae* in a 2-year nationwide surveillance study in Germany. Antimicrobial Agents Chemotherapy. **48**(11):4444-4446.
260. **Bandeira, M., P. A. Carvalho, A. Duarte, and L. Jordão.** 2014. Exploring Dangerous Connections between *Klebsiella pneumoniae* Biofilms and Healthcare-Associated Infections. Pathogens. **3**:720-731.
261. **Stepanovic, S., D. Vukovic, V. Hola, G. Di Bonaventura, S. Djukic, I. Cirkovic, and F. Ruzicka.** 2007. Quantification of biofilm in microtiter plates: overview of testing conditions and practical recommendations for assessment of biofilm production by *Staphylococci*. Apmis. **115**(8):891-899.
262. **Charlebois, A., M. Jacques, and M. Archambault.** 2014. Biofilm formation of *Clostridium perfringens* and its exposure to low-dose antimicrobials. Front Microbiol. **5**:183.
263. **Al-Fattani, M. A., and L. J. Douglas.** 2006. Biofilm matrix of *Candida albicans* and *Candida tropicalis*: chemical composition and role in drug resistance. J Med Microbiol. **55**:999-1008.
264. **Alvim, D.C.S.S., A. F. M. Ferreira, M. A. Leal, L. M. A. Oliveira, A. M. N. Botelho, A. C. N. Botelho, A. M. S. Figueiredo, S. E. L. Fracalanzza, L. M. Teixeira, and T. C. A.**

Pinto. 2019. Biofilm production and distribution of pilus variants among *Streptococcus agalactiae* isolated from human and animal sources. Biofouling. **35**:8.

265. **Azeredo, J., N. F. Azevedo, R. Briandet, N. Cerca, T. Coenye, A. R. Costa, M. Desvaux, G. Di Bonaventura, M. Hébraud, Z. Jaglic, M. Kačániová, S. Knøchel, A. Lourenço, F. Mergulhão, R. L. Meyer, G. Nychas, M. Simões, O. Tresse, and C. Sternberg.** 2017. Critical review on biofilm methods. Critic Rev Microbiol. **43**(3):313-351.

266. **Malone, M., D. M. Goeres, I. Gosbell, K. Vickery, S. Jensen, and P. Stoodley.** 2016. Approaches to biofilm-associated infections: the need for standardized and relevant biofilm methods for clinical applications. Expert Rev Anti Infect Ther. **15**(2):147–156.

267. **Marks, L.R., L. Mashburn-Warren, M. J. Federle, and A. P. Hakansson.** 2014. *Streptococcus pyogenes* biofilm growth in vitro and in vivo and its role in colonization, virulence, and genetic exchange. J Infect Dis. **210**(1):25-34.

268. **Ramstedt, M., I. A. C. Ribeiro, H. Bujdakova, F. J. M. Mergulhão, L. Jordão, P. Thomsen, M. Alm, M. Burmølle, T. Vladkova, F. Can, M. Reches, M. Riool, A. Barros, R. L. Reis, E. Meaurio, J. Kikhney, A. Moter, S. A. J. Zaat, and J. Sjollema.** 2019. Evaluating Efficacy of Antimicrobial and Antifouling Materials for Urinary Tract Medical Devices: Challenges and Recommendations. Macromol Biosci. **19**(5):e1800384.

269. **Tamura, G. S., J. M. Kuypers, S. Smith, H. Raff, and C. E. Rubens.** 1994. Adherence of group B *Streptococci* to cultured epithelial cells: roles of environmental factors and bacterial surface components. Infection and Immunity. **62**(6):2450-2458.

270. **Martins, E. R., J. Melo-Cristino, and M. Ramirez.** 2010. Evidence for Rare Capsular Switching in *Streptococcus agalactiae*. J Bacteriol. **192**(5):1361-1369.

271. **Balian, A., J. G. Gonzalez, N. Bastida, K. K. Akhtar, B. A. Borsa, and F. J. Hernandez.** 2019. Kinetic Screening of Nuclease Activity using Nucleic Acid Probes. Journal of Visualized Experiments: **153**:e60005.

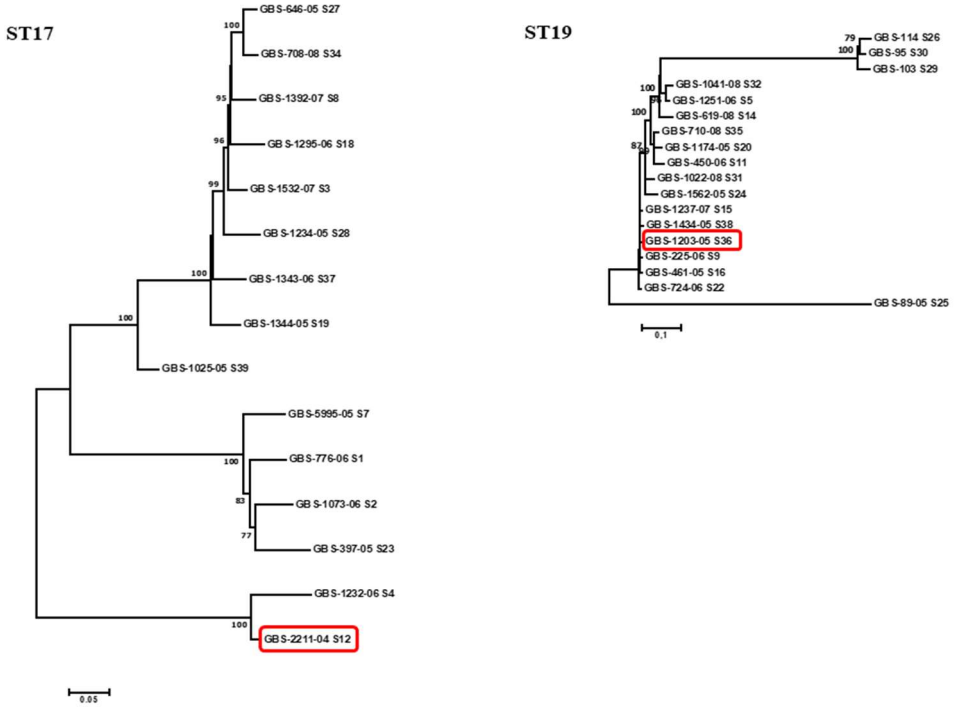
272. **Gupta, S.K., B. R. Padmanabhan, S. M. Diene, R. Lopez-Rojas, M. Kempf, L. Landraud, and J. Rolain.** 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrobial Agents Chemotherapy. **58**:212e20.

Supplementary Material

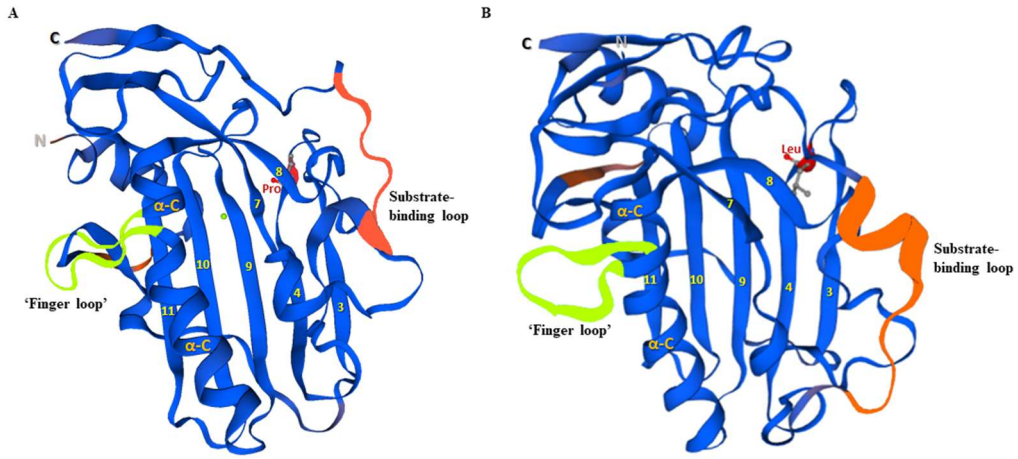
Supplementary Table S3.1 *S. agalactiae* human strains data information.

Isolate ID	Isolation Country	Year	Isolation Source	Host Gender	Host Age	Disease	ENA Accession Sample	ENA ID
1022-08	Portugal	2008	vaginal/rectal	female	33	carrier	ERS5323489	PT_GBS0001
1025-05	Portugal	2005	vaginal/rectal	female	29	carrier	ERS5323490	PT_GBS0002
103/#923	Germany	2001-2003	blood/CSF	unknown	< 3 months	invasive	ERS5323491	PT_GBS0003
1041-08	Portugal	2008	vaginal/rectal	female	38	carrier	ERS5323492	PT_GBS0004
1073-06	Portugal	2006	vaginal/rectal	female	17	carrier	ERS5323493	PT_GBS0005
114/#928	Germany	2001-2003	blood/CSF	unknown	< 3 months	invasive	ERS5323494	PT_GBS0006
1174-05	Portugal	2005	vaginal/rectal	female	22	carrier	ERS5323495	PT_GBS0007
1203-05	Portugal	2005	vaginal/rectal	female	21	carrier	ERS5323496	PT_GBS0008
1232-06	Portugal	2006	vaginal/rectal	female	22	carrier	ERS5323497	PT_GBS0009
1234-05	Portugal	2005	vaginal/rectal	female	34	carrier	ERS5323498	PT_GBS0010
1237-07	Portugal	2007	vaginal/rectal	female	29	carrier	ERS5323499	PT_GBS0011
1251-06	Portugal	2006	vaginal/rectal	female	71	carrier	ERS5323500	PT_GBS0012
1295-06	Portugal	2006	vaginal/rectal	female	35	carrier	ERS5323501	PT_GBS0013
1343-06	Portugal	2006	vaginal/rectal	female	35	carrier	ERS5323502	PT_GBS0014
1344-05	Portugal	2005	vaginal/rectal	female	20	carrier	ERS5323503	PT_GBS0015
1392-07	Portugal	2007	vaginal/rectal	female	25	carrier	ERS5323504	PT_GBS0016
1434-05	Portugal	2005	vaginal/rectal	female	31	carrier	ERS5323505	PT_GBS0017
1532-07	Portugal	2007	vaginal/rectal	female	30	carrier	ERS5323506	PT_GBS0018
1562-05	Portugal	2005	vaginal/rectal	female	20	carrier	ERS5323507	PT_GBS0019
2211-04	Angola	2004	CSF	male	< 12 years	meningitis	ERS5323508	PT_GBS0020
225-06	Portugal	2006	vaginal/rectal	female	25	carrier	ERS5323509	PT_GBS0021
397-05	Portugal	2005	vaginal/rectal	female	29	carrier	ERS5323510	PT_GBS0022
450-06	Portugal	2006	vaginal/rectal	female	24	carrier	ERS5323511	PT_GBS0023
461-05	Portugal	2005	vaginal/rectal	female	20	carrier	ERS5323512	PT_GBS0024
5995-05	Angola	2005	CSF	male	< 12 years	meningitis	ERS5323513	PT_GBS0025
619-08	Portugal	2008	vaginal/rectal	female	49	carrier	ERS5323514	PT_GBS0026
646-05	Portugal	2005	vaginal/rectal	female	33	carrier	ERS5323515	PT_GBS0027
708-08	Portugal	2008	vaginal/rectal	female	23	carrier	ERS5323516	PT_GBS0028
710-08	Portugal	2008	vaginal/rectal	female	41	carrier	ERS5323517	PT_GBS0029
724-06	Portugal	2006	vaginal/rectal	female	25	carrier	ERS5323518	PT_GBS0030
776-06	Portugal	2006	vaginal/rectal	female	59	carrier	ERS5323519	PT_GBS0031
89-05	Portugal	2005	vaginal/rectal	female	26	carrier	ERS5323520	PT_GBS0032
95/#916	Germany	2001-2003	blood/CSF	unknown	< 3 months	invasive	ERS5323521	PT_GBS0033

ID, Identification; CSF, Cerebrospinal Fluid



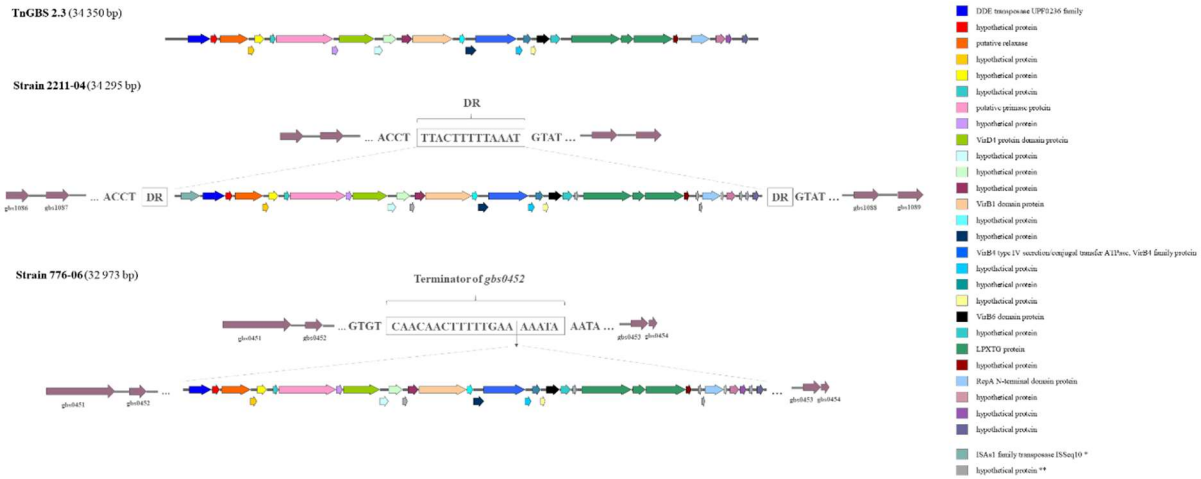
Supplementary Figure S3.1 Phylogenetic relationship of *S. agalactiae* strains within each ST. Both core-genome SNP-based phylogenetic trees were reconstructed using the Neighbor-Joining method [192] with the Maximum Composite Likelihood model [193]. Bootstraps values are shown next to the respective nodes.



Supplementary Figure S3.2 Putative structures of *S. agalactiae* NucA. For NEM316 (A) and all 33 strains under study (B), the nuclease domain of NucA, lacking the N-terminal transmembrane secretion signal sequence (checked with SignalP-5-0 Server), was used for structure assessment from SWISS-MODEL web server, using the crystal structure of NucA NEM316 as model [200]. Accordingly, the global structure of NucA NEM316 is comprised of a central anti-parallel β-sheet flanked on the 'front' face by a ββα metal-finger (ββα-Me) motif (comprised of β-strands 7-8 and α-hélix C) that outlines the compact active center, while its deepest recess is formed by the core β-sheet (β-strands 4, 9 and 10) [200]. For each structure, the proline (in NEM316) or leucine (in all 33 strains) amino acid residues (the only non-synonymous difference seen between proteins) in β-strand 4 as well as the 'finger loop' and the substrate-binding loop are shown.

Supplementary Table S3.2 List of accessory genes exclusive of DNase(-) ST-17 2211-04 strain.

Product	
Phage integrase	Hypothetical protein
Hypothetical protein	Hypothetical protein
LexA repressor	Phage protein
Transcriptional regulator	Phage terminase, small subunit
Phage antirepressor protein	Phage terminase, large subunit
Hypothetical protein	Phage minor capsid protein
Hypothetical protein	Phage minor capsid protein
Hypothetical protein	Hypothetical protein
Hypothetical protein	Hypothetical protein
Hypothetical protein	Hypothetical protein
Hypothetical protein	Hypothetical protein
Hypothetical protein	Hypothetical protein
Phage recombination protein Bet	Hypothetical protein
Phage protein	Phage protein
Hypothetical protein	Putative minor capsid protein - phage associated
Hypothetical protein	Phage minor capsid protein
Single-stranded DNA-binding protein	Phage minor capsid protein
Phage protein	Phage major tail shaft protein
Hypothetical protein	hypothetical protein
Phage protein	Phage protein
Phage protein	Hypothetical protein
Hypothetical protein	Hypothetical protein
Phage protein	Paratox
Hypothetical protein	Phage lysin, N-acetylmuramoyl-L-alanine amidase

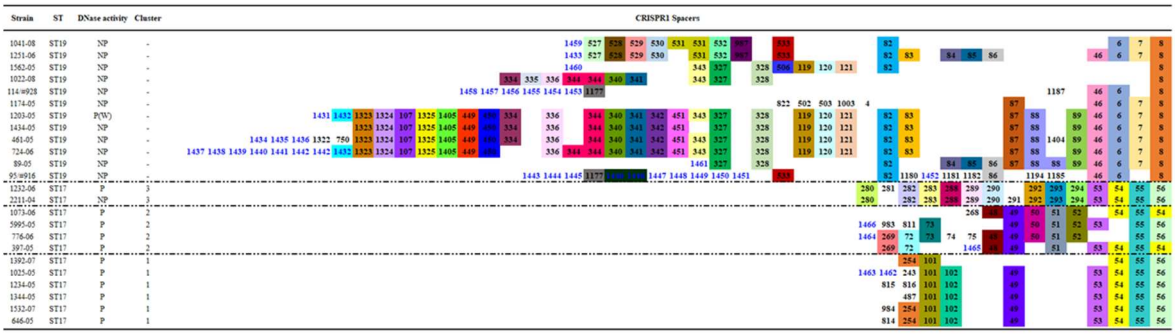


Supplementary Figure S3.3 Schematic representation of ST17 DNase(-) 2211-04 and DNase(+) 776-06 transposons. The local of insertion of the transposon for each ST17 strain and the homology to TnGBS2.3 (GenBank accession number KC492040.1) are shown. The transposons carried by DNase(-) 2211-04 and DNase(+) 776-06 are 94% and 98% homologous (99.99% and 99.97% identity) to TnGBS2.3, respectively. The ORFs of the transposons are represented by colored arrows. * represents one ORF exclusive of the transposon of ST17 DNase(-) 2211-04, while ** represents ORFs absent in transposon TnGBS2.3.

Supplementary Table S3.3 List of virulence factors differently present in DNase(+) ST19 1203-05.

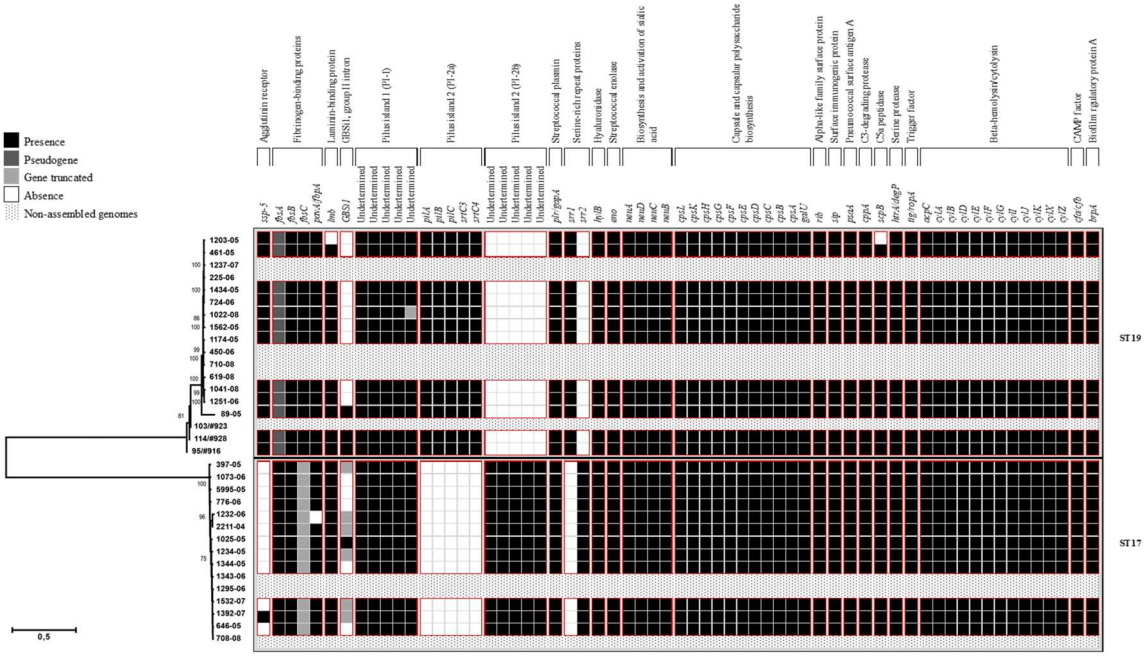
			ST17 strains		ST19 strains	
Virulence Factors	**NEM316 old locus tag [a]**	**2603V/R locus tag [b]**	**DNase(-) 2211-04**	**DNase(+) (n=14)**	**DNase(+) 1203-05**	**DNase(-) (n=17)**
Laminin-binding protein (Lmb)	gbs1307	SAG1234	√	√	---	√
Serine-rich repeat protein Srr-1	---	SAG1462	---	---	2 smaller ORFs	√
C5a peptidase	gbs1308	SAG1236	√	√	---	√

[a] ORFs designations according to NEM316 genome (GenBank accession number NC004368.1)
[b] ORFs designations according to 2603V/R genome (GenBank accession number NC004116.1)



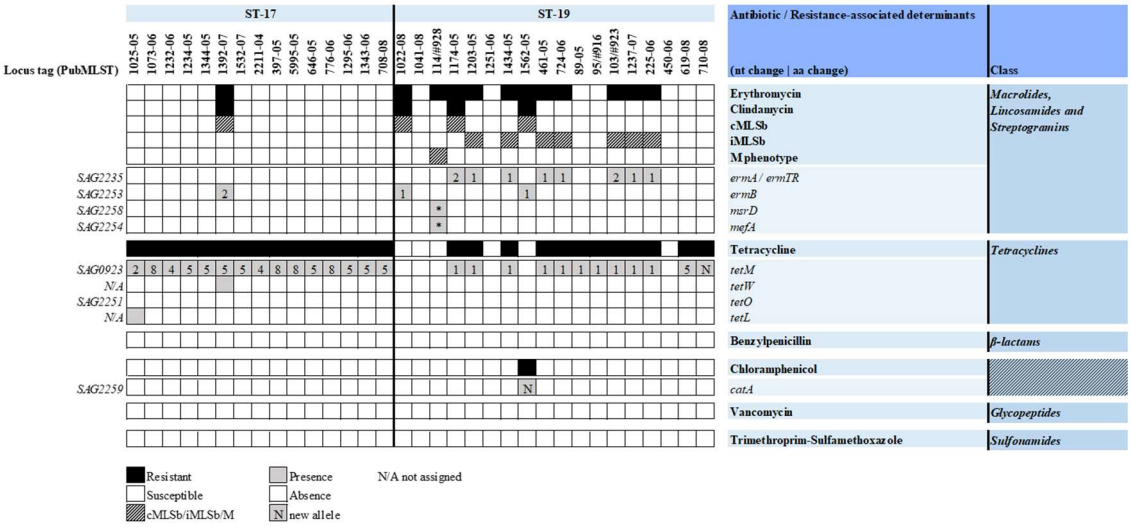
Supplementary Figure S3.4 Diversity of the CRISPR1 locus in *S. agalactiae* ST17 and ST19 strains. Spacers are represented in arbitrary colored boxes. Arrays are oriented with respect to the trailer (begin) sequence located on the right. Strains were arranged according to the CRISPR1 spacer content and grouped according to their ST. Strain names, ST and phenotype for DNase activity are given on the left. Major strain clusters previously defined for ST17 [104] are shown.

Supplementary Table S3.4 Allelic diversity of genes putatively involved in *S. agalactiae* virulence among ST17 and ST19 strains. Due to the massive extent of this table, it was impossible to present it in a printable format. Please access the following link to review the table: https://zenodo.org/record/4574483#.YD6bhWj7RPY

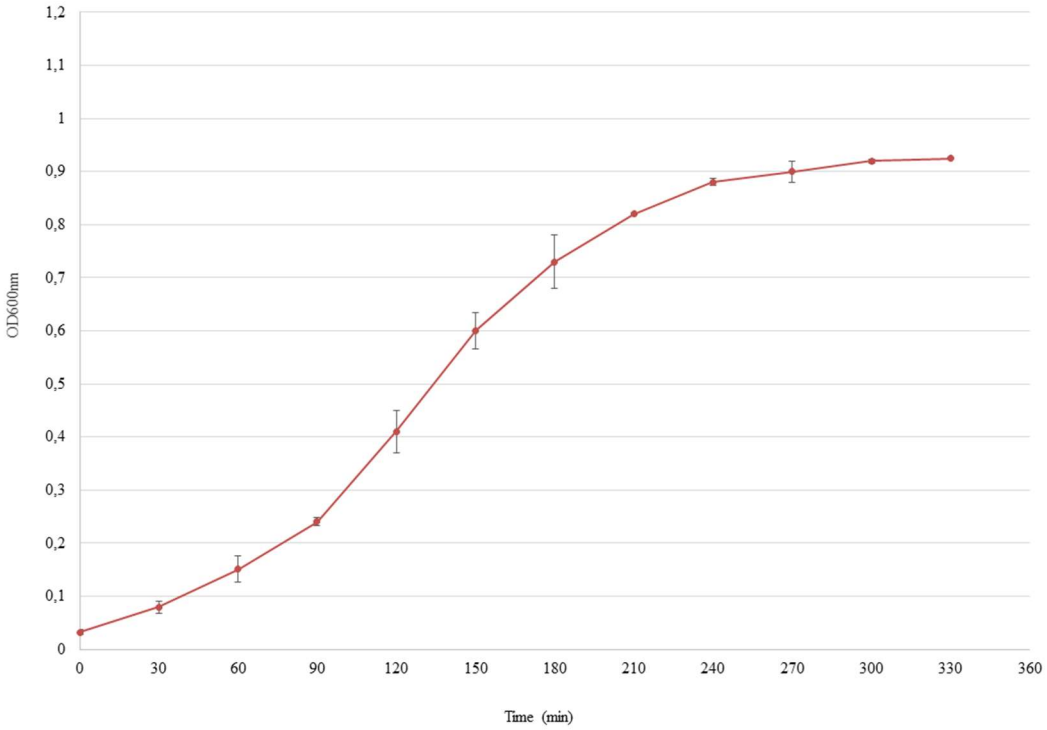


Supplementary Figure S3.5 Distribution of genes putatively involved in *S. agalactiae* virulence. The core-genome SNP-based phylogenetic tree was reconstructed using the Neighbor-Joining method [192] with the Maximum Composite Likelihood model [193]. Bootstraps values are shown next to the respective nodes. The 63 genes presented in the figure are previously identified virulence determinants and genes identified after querying draft genome sequences against the VFDB [197]. Each line refers to a specific isolate and each column to a specific gene. Gene are clustered according to their function. Black squares represent gene presence while white areas represent loci not found for specific isolates.

Supplementary Table S3.5 Antimicrobial resistance among ST17 and ST19 strains. Due to the massive extent of this table, it was impossible to present it in a printable format. Please access the following link to review the table: https://zenodo.org/record/4317773#.X9O4n9j7RPY



Supplementary Figure S3.6 Distribution of antibiotic resistance phenotypes and genes in ST17 and ST19 *S. agalactiae* strains. Antimicrobial susceptibility testing was performed by E-test® (minimum inhibitory concentration (MIC) determination) or disk diffusion method according to European Committee for Antimicrobial and Susceptibility Testing (EUCAST) recommendations using Mueller Hinton agar supplemented with 5% defibrinated horse blood and 20 mg/liter of β-NAD (MH-F) (Oxoid, Basingstoke, UK). The inducible clindamycin resistance phenotype was identified by the double-disk diffusion method. EUCAST breakpoints (V10.0; January 2020) (http://www.eucast.org/clinical_breakpoints/) were used to classify strains as susceptible or resistant. For each strain, the presence of antimicrobial resistance genes was search by querying the draft genome assembly against the following publicly available reference databases using ABRIcate v0.8 (https://github.com/tseemann/abricate): CARD (https://card.mcmaster.ca/), ResFinder (http://www.genomicepidemiology.org/), ARG-ANNOT (http://backup.mediterranee-infection.com/article.php?laref=282&titre=arg-annot) [272] and NCBI AMRFinderPlus (http://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/). For comparative purposes, analyses were also performed using trimmed reads whenever that was an option. The presence/absence and/or variability of each putative identified hit was confirmed by an assembly-free strategy using Snippy 4.6 software (https://github.com/tseemann/snippy), while BLASTp against the nr protein sequences database was used to confirm protein functionality. The identified hits were then submitted at to PubMLST online platform (http://pubmlst.org/) for allele determination. The predicted results were compared with antimicrobial susceptibility testing results. Resistance is represented by black squares, while susceptibility in white squares. The gene determinant of resistance is represented in grey light squares for each resistant strain, and the different alleles found are represented in numbers (whenever it was possible to determine). cMLSB, constitutive macrolide-lincosamide-streptogramin B resistance phenotype; iMLSB, inducible macrolide-lincosamide-streptogramin B resistance; M, macrolide resistance



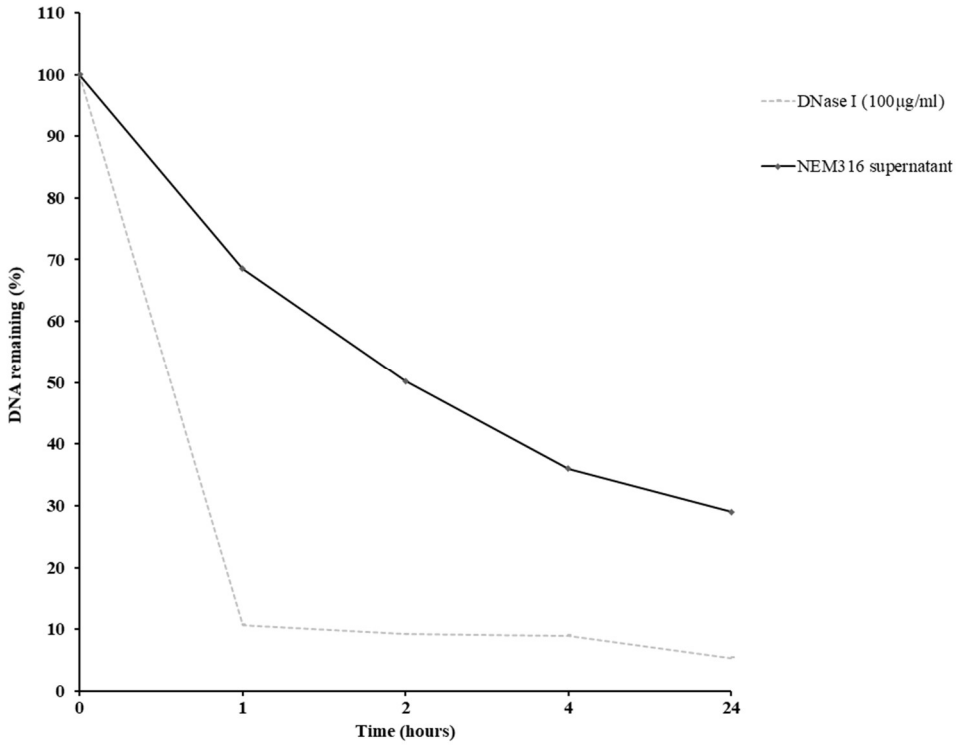
Supplementary Figure S4.1 Growth curve of *S. agalactiae* strain NEM316 at 37ºC. Bacteria were inoculated into THB from an overnight culture and growth was monitored by measuring the OD_{600}. Three independent experiments were performed.

Supplementary Table S4.1 List of genes for which expression values were obtained through RNA-seq for NEM316 at exponential growth phase. Due to the massive extent of this table, it was impossible to present it in a printable format. Please access the following link to review the table: https://zenodo.org/record/4243111#.X6Hpooj7RPY

Supplementary Table S5.1 Characterization of *S. agalactiae* clinical strains.

Strain	Origin	Carrier/ Infection	Capsular serotype/genotype	Sequence Type	DNase activity	Biofilm formation with various media (48h incubation)		
						THB	THB+1% Glucose	RPMI-glucose+15%IHP
2211-04	Human	I	III-2		NP	WBP	WBP	SBP
397-05	Human	C	III-2		P	WBP	MBP	SBP
646-05	Human	C	III-2		P	WBP	WBP	SBP
1025-05	Human	C	III-2		P	WBP	MBP	SBP
5590-05	Human	I	III-2		P	WBP	WBP	SBP
5659-05	Human	I	III-2		P	WBP	WBP	SBP
5995-05	Human	I	III-2		P	WBP	WBP	SBP
1234-05	Human	C	III-2		P	MBP	SBP	SBP
1344-05	Human	C	III-2	ST17	P	WBP	MBP	SBP
466-06	Human	C	III-2		P	WBP	MBP	SBP
776-06	Human	C	III-2		P	WBP	MBP	SBP
1073-06	Human	C	III-2		P	WBP	MBP	SBP
1232-06	Human	C	III-2		P	WBP	MBP	MBP
1295-06	Human	C	III-2		P	WBP	WBP	WBP
1343-06	Human	C	III-2		P	WBP	MBP	SBP
1392-07	Human	C	III-2		P	MBP	SBP	SBP
1532-07	Human	C	III-2		P	MBP	SBP	SBP
708-08	Human	C	III-2		P	WBP	WBP	SBP
89-05	Human	C	V		NP	WBP	WBP	WBP
461-05	Human	C	III-1		NP	WBP	WBP	SBP
713-05	Human	C	III-1		NP	WBP	WBP	SBP
1174-05	Human	C	III-1		NP	WBP	WBP	SBP
1203-05	Human	C	III-1		P	WBP	WBP	SBP
1434-05	Human	C	III-1		NP	WBP	WBP	SBP
1562-05	Human	C	III-1		NP	WBP	WBP	SBP
225-06	Human	C	III-1		NP	WBP	WBP	SBP
450-06	Human	C	III-1		NP	WBP	MBP	SBP
724-06	Human	C	III-1	ST19	NP	WBP	MBP	SBP
1251-06	Human	C	III-1		NP	WBP	WBP	SBP
1237-07	Human	C	III-1		NP	WBP	MBP	SBP
619-08	Human	C	III-1		NP	WBP	WBP	SBP
710-08	Human	C	III-1		NP	WBP	MBP	SBP
1022-08	Human	C	III-1		NP	WBP	WBP	SBP
1041-08	Human	C	III-1		NP	WBP	WBP	SBP
95/#916	Human	I	II		NP	WBP	MBP	SBP
103/#923	Human	I	II		NP	WBP	MBP	SBP
114/#928	Human	I	II		NP	WBP	MBP	SBP
211/#960	Human	I	II		NP	WBP	MBP	SBP

C, Carrier; I, Infection; P, Producer; NP, Non-producer; THB, Todd Hewitt Broth; IHP, Inactivated Human Plasma; WBP, Weak Biofilm Producer; MBP, Moderate Biofilm Producer; SBP, Strong Biofilm Producer



Supplementary Figure S5.1 Quantitative DNase assays displaying differential DNase activity between commercial DNase I and culture supernatant from stationary phase of reference strain NEM 316.
1 µg of DNA (amplicon *atr*) incubated with DNase I (100 µg/ml) (Qiagen, Germantown, United States) or *S. agalactiae* culture supernatant for 1h, 2h, 4h, overnight at 37°C. Fluorescent PicoGreen dye (Invitrogen) was used to quantify the dsDNA.