UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA

Lancefield group C and G streptococci in human infection: Molecular typing, virulence and antimicrobial susceptibility

MARCOS DANIEL CAETANO BORGES DE PINHO

DOUTORAMENTO EM CIÊNCIAS E TECNOLOGIAS DA SAÚDE

ESPECIALIDADE MICROBIOLOGIA

2014

UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA

Lancefield group C and G streptococci in human infection: Molecular typing, virulence and antimicrobial susceptibility

MARCOS DANIEL CAETANO BORGES DE PINHO

Tese orientada pelo

PROFESSOR DOUTOR MÁRIO NUNO RAMOS D' ALMEIDA RAMIREZ

DOUTORAMENTO EM CIÊNCIAS E TECNOLOGIAS DA SAÚDE

ESPECIALIDADE MICROBIOLOGIA

A impressão desta dissertação foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 18 de Fevereiro de 2014.

ACKNOWLEDGMENTS

My first acknowledgment is to Professor Mário Ramirez, the supervisor of this thesis, for all the guidance provided and the enthusiasm he showed throughout the development of this work. I am also deeply grateful to Professor José Melo Cristino, head of Instituto de Microbiologia, Faculdade de Medicina de Lisboa, for giving me the opportunity to be part of this work team and for the active role he had in all the work included in this thesis. I also wish to thank both for all the knowledge they shared with me throughout the years, and for all the unconditional support which was far beyond the professional issues.

My acknowledgment to Prof. David McMillan, from the Queensland Institute of Medical Research and Griffith Medical Research College, Australia, Prof. Debra Bessen, from the New York Medical College, United States of America, Prof. Constança Pomba, from the Universidade de Lisboa, Portugal, Prof. Silvia Preziuso, from the University of Camerino, Italy, Prof. Antina Lübke Becker and Prof. Lothar Heinz Wieler, from the Freie Universität Berlin, for their collaboration and invaluable contribution to this work, which allowed the purposes of this thesis to be broadened.

I also wish to thank all my colleagues at Instituto de Microbiologia and Unidade de Microbiologia Molecular e Infecção for the nice working environment. A special thanks to Sandra Matos for her help in characterizing some of the bacterial isolates studied. I am also grateful to Ana Friães, Catarina Costa, Sandra Aguiar and João Carriço for their friendship and encouragement.

I thank to my closest family and friends for their care and constant motivation. I express all my gratitude to my parents, whom support and love through all phases of my life have made possible to reach this point. To my sister and brother, Mónica and Miguel, to Bruno, to my grandparents, to Tetyana, Leonid and Olga, thank you for all the support and patience.

To Olena, Carolina and Rafael, thank you for being part of me.

SUMMARY

Keywords: *S. dysgalactiae* subsp. *equisimilis*, *S. canis*, population, recombination.

Beta-hemolytic, large-colony-forming (diameter greater than 0.5 mm) Lancefield group C and group G streptococci (GCGS) is a group within the *Streptococcus* genus which includes several species recognized as either colonizers or pathogens in humans and animals. *Streptococcus dysgalactiae* subsp. *equisimilis* (SDE), which can express any of these two Lancefield group antigens, is the GCGS species most commonly reported in human infection worldwide and may cause a number of potentially life-threatening infections. SDE is increasingly regarded as an emerging global pathogen and is able to colonize and infect humans, while other GCGS species, such as *Streptococcus canis*, are mainly animal pathogens that occasionally infect the human host. The rising number of human infections reported to be caused by GCGS warrants a better study of their epidemiology, in order to establish the relevance of each species, and clarifying the clonal dynamics and the intra-specific factors influencing the virulence of the strains.

This work aimed at determining the GCGS species responsible for human infection in Portugal and, by assessing the genetic diversity of the isolates recovered, to define the clonal structure of this population. Thus, a special emphasis on molecular typing techniques was given, including those more commonly used to type these streptococci: pulsed-field gel electrophoresis (PFGE), *emm* typing and multilocus sequence typing (MLST). The application of these techniques to characterize collections of SDE and *S. canis* isolates recovered from other geographic locations and, for the latter species, from animal hosts, allowed to elucidate their respective population structures and provided new insights into the biology and evolution of GCGS.

Initially, the speciation and characterization of GCGS isolates recovered from human infections in Portugal identified the central role of SDE. Not only was there a weak correlation between *emm* typing and PFGE results, but the polyclonal origin of the SDE population in this region was revealed by each of the methods used that generated multiple partitions. Furthermore, a correlation between *emm* type and invasive disease potential was suggested.

A more global snapshot of the clonal architecture of this pathogen population was achieved by developing a MLST scheme which was applied to an expanded collection of SDE isolates recovered from distinct continents. An association of Lancefield groups with distinct MLST partitions was found and the high prevalence of a small number of widely distributed MLST sequence types (STs) suggested that a few genetic lineages dominate among SDE causing human infection worldwide. The occurrence of intra-specific and inter-specific recombination with *Streptococcus pyogenes* (Lancefield group A streptococcus, [GAS]), involving the housekeeping genes used in MLST was detected.

The poor correlation between *emm* typing and either PFGE or MLST defined groups was illustrated by lineages displaying distinct *emm* types and the presence of the same *emm* type in unrelated genetic backgrounds, as defined by both techniques. These observations suggested the existence of recombinational replacements involving the *emm locus* and question the value of *emm* typing to accurately ascertain the genetic relatedness of SDE strains.

The characterization of the antimicrobial susceptibility patterns presented by SDE found a high proportion of levofloxacin-resistant isolates (12%) associated with multiple genetic lineages. Sequence analysis of the quinolone resistance-determining regions of the *gyrA* and *parC* genes of representative resistant and susceptible isolates showed that full resistance was associated with mutations in both GyrA and ParC. As observed for the housekeeping genes used in MLST, recombination with GAS DNA in some *parC* alleles was evident, though this phenomenon was not exclusively associated with resistance.

The final part of the work in this thesis focused on *S. canis*, the second most frequent GCGS species isolated from human infections in Portugal. A collection of *S. canis* isolates recovered from infections in both humans and animals, collected in Portugal and abroad, were characterized by employing the same typing methods used for SDE. The *S. canis* population was polyclonal, and several genetic lineages were shown to possess the ability to infect the human host. The zoonotic nature of *S. canis* infection was demonstrated, as identical genetic lineages were found infecting house pets and humans, indicating that they constitute a single population. Phylogenetic analysis showed that *S. canis* was a divergent *taxon* of SDE and GAS and unveiled the acquisition of genetic material of SDE by *S. canis*. The presence of *emm*-like genes was

restricted to a few *S. canis* isolates and correlated with some MLST-based genetic lineages.

Globally, this thesis contributes to the current knowledge of the molecular epidemiology and evolutionary relationships among members of the two GCGS species studied. The clonal relationships among strains were elucidated and MLST schemes for SDE and *S. canis* were established, providing useful tools for future studies of their population dynamics. The use of *emm* typing was shown to be complemented by applying other typing methods and the role of the M protein in SDE virulence was reinforced. The evidence found for recombinational replacements between SDE and GAS in several *loci* and, to a smaller extent, between SDE and *S. canis*, indicates that horizontal gene transfer events are important mechanisms driving genetic variability in GCGS populations which may impact key bacterial functions such as virulence and antimicrobial resistance.

RESUMO

Palavras-chave: *S. dysgalactiae* subsp. *equisimilis*, *S. canis*, população, recombinação.

Os estreptococos beta-hemolíticos formadores de colónias grandes (diâmetro superior a 0.5 mm) dos grupos C e G de Lancefield (GCGS) formam um grupo dentro do género *Streptococcus* que inclui várias espécies reconhecidas como colonizadoras ou agentes patogénicos no Homem e em animais. *Streptococcus dysgalactiae* subsp. *equisimilis* (SDE), que pode integrar cada um destes dois grupos de Lancefield, é a espécie de GCGS mais vezes reportada em infecção humana a nível mundial e pode causar diversas infecções potencialmente fatais. SDE é cada vez mais visto como um agente patogénico emergente, com a capacidade de colonizar e infectar o Homem, enquanto as outras espécies de GCGS, tais como *Streptococcus canis*, são principalmente agentes patogénicos de animais que ocasionalmente infectam o hospedeiro humano. O número crescente de infecções humanas atribuídas aos GCGS exige um estudo reforçado da sua epidemiologia, de modo a estabelecer a relevância de cada espécie e permitindo que a dinâmica clonal e os factores intra-específicos que influenciam a virulência das estirpes sejam clarificados.

Este trabalho teve como objectivo identificar as espécies de GCGS responsáveis por infecção humana em Portugal e, através da avaliação da diversidade genética das estirpes recolhidas, definir a estrutura clonal da população. Deste modo, foi dado especial ênfase às técnicas de tipagem molecular, incluindo aquelas que mais habitualmente são utilizadas para a tipagem destes estreptococos: a electroforese em campo pulsado (PFGE), a tipagem *emm* e "multilocus sequence typing" (MLST). A aplicação destas técnicas para caracterizar colecções de estirpes de SDE e *S. canis* isoladas noutras regiões geográficas e, no caso da última espécie, de hospedeiros animais, permitiu elucidar as respectivas estruturas populacionais e percepcionar novos aspectos da biologia e evolução dos GCGS.

Inicialmente, a especiação e caracterização de estirpes de GCGS isoladas de infecções humanas em Portugal mostrou o papel central de SDE. Verificou-se uma fraca correlação entre os resultados obtidos pela tipagem *emm* e PFGE e a aplicação destas técnicas revelou a origem policlonal da população de SDE nesta região, dado que cada

um dos métodos utilizados gerou múltiplas divisões. Adicionalmente, os resultados sugeriram a existência de uma correlação entre o tipo *emm* e o potencial de doença invasiva.

Uma imagem mais geral da estrutura apresentada pela população deste agente patogénico foi alcançada através do desenvolvimento de um esquema de MLST, aplicado a um conjunto alargado de estirpes de SDE isoladas em continentes distintos. Esta análise permitiu estabelecer uma associação entre grupos de Lancefield e divisões distintas de MLST. Foi possível observar uma elevada prevalência de um pequeno número de "sequence types" (STs) na população. A ampla distribuição geográfica destes STs sugeriu o predomínio de algumas linhagens genéticas entre as estirpes de SDE que causam infecção humana em todo o mundo. Foi detectada ainda a ocorrência de recombinação intra-específica e inter-específica com *Streptococcus pyogenes* (estreptococos do grupo A de Lancefield, [GAS]), envolvendo genes "housekeeping" utilizados no MLST.

A fraca correlação entre tipos *emm* e os grupos definidos por PFGE ou por MLST foi ilustrada pela existência de linhagens com tipos *emm* distintos e pela presença do mesmo tipo *emm* em patrimónios genéticos não relacionados, tal como definido pelas duas técnicas de tipagem. Estas observações sugeriram a existência de substituições recombinacionais envolvendo este *locus* e questionam o valor da tipagem *emm* para determinar correctamente a relação genética de estirpes de SDE.

A caracterização dos perfis de susceptibilidade aos antimicrobianos apresentados por SDE permitiu detectar uma elevada percentagem de estirpes resistentes à levofloxacina (12%), associada a múltiplas linhagens genéticas. A análise das sequências das regiões determinantes da resistência às quinolonas nos genes *gyrA* e *parC* de estirpes resistentes e susceptíveis representativas, mostrou que a aquisição de resistência se encontra associada a alterações dos aminoácidos das proteínas codificadas pelos dois genes. A existência de recombinação com ADN de GAS foi evidente em alguns alelos do gene *parC*, como observado para os genes "housekeeping" utilizados no MLST. No entanto, este acontecimento não foi associado exclusivamente com o desenvolvimento de resistência.

A parte final do trabalho descrito nesta tese focou-se na segunda espécie de GCGS mais frequentemente isolada de infecções humanas em Portugal, *S. canis*. Uma colecção constituída por estirpes de *S. canis* isoladas de infecções em seres humanos e animais,

recolhidas em Portugal e no estrangeiro, foi caracterizada recorrendo aos mesmos métodos de tipagem utilizados para SDE. A policlonalidade da população de *S. canis* foi evidenciada e várias linhagens genéticas mostraram ter a capacidade de infectar o hospedeiro humano. A natureza zoonótica da infecção por *S. canis* foi demonstrada pois as mesmas linhagens genéticas foram encontradas a infectar animais domésticos e seres humanos, indicando a existência de uma única população. A análise filogenética mostrou que *S. canis* é um *taxon* divergente das espécies SDE e GAS e revelou a aquisição de material genético de SDE por *S. canis*. A presença de genes similares ao gene *emm* restringiu-se a uma pequena proporção das estirpes de *S. canis* e correlacionou-se com algumas das linhagens genéticas definidas por MLST.

Globalmente, este trabalho contribuiu para o conhecimento actual da epidemiologia molecular e das relações evolutivas entre os membros das duas espécies de GCGS estudadas. As relações clonais entre as estirpes foram elucidadas e esquemas de MLST foram estabelecidos para SDE e *S. canis*, constituindo ferramentas úteis para futuros estudos sobre a dinâmica populacional destas espécies. Mostrou-se que a utilização da tipagem *emm* é complementada pela aplicação de outros métodos de tipagem e a importância da proteína M na virulência de SDE foi reforçada. As indicações encontradas da ocorrência de trocas recombinacionais entre SDE e GAS em vários *loci* e, em menor extensão, entre SDE e *S. canis*, indica que os eventos de transferência genética horizontal são mecanismos motrizes da variabilidade genética em populações de GCGS, podendo afectar aspectos fulcrais da biologia destas espécies, como a virulência e a resistência aos antimicrobianos.

THESIS OUTLINE

The main purpose of the current work was to provide insights into the molecular epidemiology of GCGS causing human infections.

Chapter 1 is a general introduction describing the main aspects of GCGS biology. As the current thesis focuses on human infections by these bacteria, a special emphasis is given to SDE, the species most commonly found and the subject of most of the laboratory work described in this thesis. A few aspects regarding the biology of other GCGS species are also studied, especially those which have also been isolated from the human host. A more detailed description of the typing methodologies used is also found under this section. This chapter concludes by stating the aims of the current work.

Chapter 2 describes the epidemiology of GCGS recovered from invasive and non-invasive infections in Portugal, and involves the characterization of 116 isolates recovered from 1998 to 2004 in hospital associated laboratories across the country. It starts by identifying the species involved and, since only SDE is detected, it proceeds with a detailed characterization by *emm* typing and PFGE of the isolates of this species. Also, the association of specific *emm* types and PFGE defined clones with invasive disease potential is statistically evaluated. The epidemiologic characterization of SDE presented in this chapter is supplemented by the data described in chapters 3 and 4.

Chapter 3 centers on the development of an MLST scheme for the characterization of SDE. In collaboration with groups from Australia and the United States, an epidemiological and genetically diverse collection of isolates was obtained, including 36 isolates recovered in Portugal and representative of several *emm* types encountered in the work reported in chapter 2. The application of the MLST technique allowed a global view of the SDE population causing infections in humans and to estimate the occurrence of recombination involving the housekeeping genes used in the MLST scheme.

Chapter 4 describes the antimicrobial resistance patterns observed among an expanded collection of SDE human isolates recovered in Portugal (n = 314, recovered from 1998 to 2005), with a special focus on levofloxacin resistance. Prompted by the high resistance rate found and by previous suggestions that members of the *S. dysgalactiae taxon* could mediate the appearance of resistance in GAS through horizontal gene transfer events, the molecular mechanisms of fluoroquinolone resistance

in SDE were investigated and phylogenetic analyses were conducted with the sequence data generated from the resistance-determining regions of the relevant genes.

Chapter 5 focuses on a distinct Lancefield group G species, *S. canis*, which was found as an infrequent agent of human infection in the last years of the study. This chapter reports the molecular characterization of 85 *S. canis* isolates employing the same typing techniques used for SDE, as described in the previous chapters. Special emphasis was given to the development of a universally applicable MLST scheme for *S. canis*. Given the low number of human infections caused by this species, the collection was expanded by adding isolates recovered from animal sources, mainly house pets, which allowed the characterization of the *S. canis* population and provided reinforced evidence of the zoonotic nature of human infection.

Chapter 6 is a general discussion which aims to summarize and articulate the main findings of the thesis. The contribution of this work is contextualized in the current issues on GCGS biology and the working questions that can be pursued in the future are highlighted.

Chapters 2 to 5 can be read independently and are reproductions of the publications listed in the next page. Only minor alterations were made relative to the published version, with the purpose of standardizing text with the other contents of the thesis.

PUBLICATIONS INCLUDED IN THIS THESIS

Chapter 2:

Pinho MD, Melo-Cristino J, Ramirez M. 2006. Clonal relationships between invasive and noninvasive Lancefield group C and G streptococci and *emm*-specific differences in invasiveness. J. Clin. Microbiol. **44**:841-846.

Chapter 3:

McMillan DJ, Bessen DE, Pinho M, Ford C, Hall GS, Melo-Cristino J, Ramirez M. 2010. Population genetics of *Streptococcus dysgalactiae* subspecies *equisimilis* reveals widely dispersed clones and extensive recombination. PLoS One **23**;5:e11741.

Chapter 4:

Pinho MD, Melo-Cristino J, Ramirez M. 2010. Fluoroquinolone resistance in *Streptococcus dysgalactiae* subsp. *equisimilis* and evidence for a shared global gene pool with *Streptococcus pyogenes*. Antimicrob. Agents Chemother. **54**:1769-1777.

Chapter 5:

Pinho MD, Matos SC, Pomba C, Lübke-Becker A, Wieler LH, Preziuso S, Melo-Cristino J, Ramirez M. 2013. Multilocus sequence analysis of *Streptococcus canis* confirms the zoonotic origin of human infections and reveals genetic exchange with *Streptococcus dysgalactiae* subsp. *equisimilis*. J. Clin. Microbiol. **51**:1099-1109.

ABBREVIATIONS

AW – Adjusted Wallace coefficient

CC – clonal complex[1]

CDC – Centers for Disease Control and Prevention

CI – confidence interval

CLSI – Clinical and Laboratory Standards Institute

DLV – double-locus-variant[1]

FCT – fibronectin-binding, collagen-binding and T antigen

FQ – fluoroquinolone

GAS – Lancefield group A streptococcus, *Streptococcus pyogenes*

GBS – Lancefield group B streptococcus, *Streptococcus agalactiae*

GCGS – beta-hemolytic, large-colony-forming (diameter, >0.5 mm) Lancefield group C and group G streptococci[2]

GCS – Lancefield group C streptococci

GGS – Lancefield group G streptococci

MIC – minimum inhibitory concentration

MLS_B – macrolide, lincosamide and streptogramin B

MLST – multilocus sequence typing

NJ – Neighbor joining

OR – odds ratio

PFGE – pulsed-field gel electrophoresis

QRDR – quinolone resistance-determining region

RDP – Recombination Detection Program

SDE – *Streptococcus dysgalactiae* subsp. *equisimilis*

SID - Simpson Index of Diversity

SLV – single-locus-variant[1]

ST – sequence type[1]

---

[1] These terms are always used in the context of the multilocus sequence typing technique.

[2] The designation GCGS used in the current work refers exclusively to large-colony-forming species. Some authors use the GCGS acronym in a broader sense which also includes the small-colony-forming (diameter, <0.5 mm) species belonging to the "*Streptococcus milleri*" or "*Streptococcus anginosus*" group. Although some beta-hemolytic "*S. milleri*" isolates can be found in the same ecologic niches of GCGS and may bear the Lancefield group C and G antigens, they constitute a genetically distinct group, differing from GCGS in the spectrum of clinical manifestations caused, in their pathogenicity and affecting different patient populations.

STSS – streptococcal toxic shock syndrome

TLV – triple-locus-variant[1]

UPGMA – unweighted-pair group method with arithmetic means

W – Wallace coefficient

TABLE OF CONTENTS

CHAPTER 6

CHAPTER 1

General Introduction

INTRODUCTORY NOTE

Beta-hemolytic, large-colony-forming (diameter greater than 0.5 mm) Lancefield group C and group G streptococci (GCGS) is a group within the *Streptococcus* genus including several bacterial species which colonize and/or infect humans and animals. Since the original description of streptococci bearing these Lancefield groups by Rebecca Lancefield in the 1930's, who found them infecting animals (1) and colonizing the vagina of parturient women (2), GCGS were thought to be essentially animal pathogens showing low pathogenicity towards the human host, in which they were present as part of the commensal microbiota (3). Moreover, the low numbers of human infections reported for GCGS, particularly when compared to other beta-hemolytic pathogens, such as *Streptococcus pyogenes* (Lancefield group A streptococcus [GAS]) or *Streptococcus agalactiae* (Lancefield group B streptococcus [GBS]), further strengthened that view and contributed to the fact that GCGS infections in humans would be overlooked in subsequent years.

A trend of increasing detection of GCGS in human infections has been noted from the late 1970's and 80's onwards in various locations around the world (3). Many of the reports published since then have described invasive infections by GCGS (4–7) and these microorganisms were implicated in triggering some of the most severe clinical syndromes caused by beta-hemolytic streptococci (8, 9). Occasionally, increased incidence of GCGS infection has been reported (10–12) and a few studies showed that the disease burden attributable to these bacteria could be similar or even higher than that of GAS or GBS (11–13). The characterization of the genome of GCGS undertaken in the last five years (14–16) confirmed their genetic similarity to other beta-hemolytic streptococci and further showed that these bacteria have a wide array of genes encoding factors that could mediate their virulence. Taken together, these observations led to the recognition of the pathogenic potential of these microorganisms towards humans, and highlighted the need for a better characterization of the GCGS strains causing infection in the human host.

The speciation of GCGS was frequently neglected over the years. Although it was long recognized that the term GCGS included several groups that could be potentially distinguished by their phenotypic traits and host distribution (3), many of the available studies in the literature did not report isolate characteristics other than the Lancefield group. Nowadays, it is recognized that the distinct taxonomic entities included in the

group differ in their significance as human pathogens. A single *taxon*, *Streptococcus dysgalactiae* subsp. *equisimilis* (SDE), is found established in the human host, being the only GCGS for which colonization is currently recognized and the most frequently detected in infection in many parts of the world (17–19). Therefore, many authors refer to GCGS isolated from human specimens as SDE (20). The spectrum of infections caused by SDE is similar to that caused by GAS and both bacteria share many proteins that were shown to be virulence factors (6, 15, 20). However, a few other GCGS species may also be found in human infections, as is the case of *Streptococcus equi* subsp. *zooepidemicus* (5) and *Streptococcus canis* (21). These are primarily found as animal pathogens and they have been isolated from human infections less frequently than SDE.

The current work focuses on the importance of GCGS for human infections and the literature review carried out in this chapter highlights the aspects of the biology of GCGS relevant for the pathogenesis in the human host. Thus, emphasis is given to SDE and to studies which unambiguously identify the isolates included. The comparison between SDE and the other GCGS species is made whenever there are differences that justify it. Particular attention is also given to *S. canis*, as it was the second most frequent GCGS species isolated in Portugal and part of the work is dedicated to it (Chapter 5). Many studies have focused on the pathogenesis of GCGS in animals, mainly regarding members of the *S. equi taxon*. Such contributions are not discussed since they are beyond the scope of this thesis, but may be referred to in relevant points.

A few issues should also be noted on the possible limitations of the available GCGS literature. In addition to the studies indicating solely the Lancefield group of the isolates, many authors characterize Lancefield group C and group G strains independently and studies that focus on either one or the other are available. In the context of the human host, it is now clear that most GCGS are SDE and, for this species, there is no evidence supporting a separation between strains bearing Lancefield group C and G antigens based on their clinical relevance. Thus, reports focusing solely on group C or group G SDE strains are included and no separate analysis is made. Finally, in a significant number of studies it is unclear whether the authors discriminate between GCGS and the small-colony-forming "*Streptococcus milleri"* strains, which can be relevant in particular ecologic niches where the two groups coexist (as is the case of the human oropharynx). The aspects mentioned above limit the value of some studies, since the relative importance of each GCGS species cannot be accurately

assessed, and these aspects were taken into consideration when conducting the literature review.

1. CLASSIFICATION OF GCGS

1.1. Morphologic and general characteristics

As members of the *Streptococcus* genus, GCGS are facultative anaerobic and catalase negative Gram-positive cocci, with spherical or ovoid cells less than 2 μm in diameter that display chains upon Gram staining (22). The guanine-cytosine content of GCGS species is within the range of 39 to 42%, as evaluated by the currently available genomes (http://www.ncbi.nlm.nih.gov/genome). GCGS are nutritionally fastidious bacteria and their growth in enriched media is enhanced by the addition of blood or serum (22). GCGS have been traditionally classified based on a set of phenotypic characteristics, including the morphology and size of the colonies, the type of hemolysis and the serologic specificity of the cell wall Lancefield group carbohydrate, which determines the classification of streptococcal isolates in this group.

The use of the hemolytic activity to classify streptococci started in the beginning of the twentieth century and is credited to Schottmüller (23). Almost all GCGS strains are beta-hemolytic, *i.e.* the colonies formed after growth in solid medium containing blood are surrounded by a zone of complete erythrocyte lysis due to the action of an exotoxin (24) (Fig. 1.1). Thus, GCGS are members of the beta-hemolytic or pyogenic group (23), which includes major human pathogens such as GAS and GBS to which GCGS are genetically related (25, 26). Some GCGS strains present variable hemolytic activity, since they can be alpha-hemolytic (partial hemolysis, due to reduction of the hemoglobin in erythrocytes which originates a green colored halo) or gamma-hemolytic (absence of hemolysis) (Fig. 1.1). Such features are described for members of the *taxon Streptococcus dysgalactiae* (27, 28) but are not exclusive of GCGS as occasional non-hemolytic isolates from other beta-hemolytic streptococci have been reported (29).

Most beta-hemolytic streptococci can be further classified serologically into distinct Lancefield groups, according to the precipitin reaction scheme proposed by Rebecca Lancefield in 1933 (1). Although Hitchcock was the first to describe the presence of a soluble specific substance on the cell wall of these streptococci (30), only after the work of Rebecca Lancefield was the polysaccharidic and group-specific nature of this substance unveiled, which was named carbohydrate C (1). The major constituents of carbohydrate C were identified as rhamnose and one amino sugar (31). Changes in content of this last component originate differences in the antigenic specificity between

bacteria of the various Lancefield groups, which is the basis of this serological technique (31).

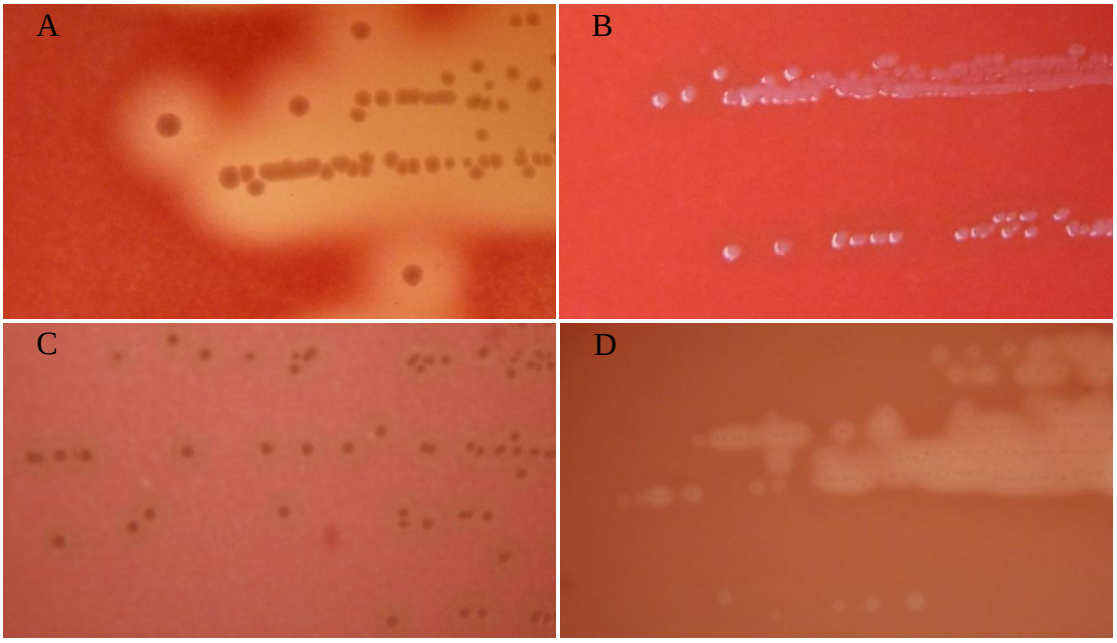

FIGURE 1.1. Colony morphology and hemolytic reaction of Lancefield group C and G bearing strains

Colonies formed after growth for 24 hours at 37ºC in tryptic soy agar supplemented with 5% (vol/vol) defibrinated sheep blood. (A) Large-colony-forming (>0.5mm) SDE isolate displaying beta-hemolysis; (B) Large-colony-forming gamma-hemolytic (no hemolysis) *Streptococcus dysgalactiae* subsp. *dysgalactiae* isolate; (C) Large-colony-forming alfa-hemolytic *S. dysgalactiae* subsp. *dysgalactiae* isolate; (D) Small-colony-forming (<0.5mm) "*S. milleri*" isolate displaying beta-hemolysis.

The Lancefield grouping scheme was a critical contribution to the classification of beta-hemolytic streptococci and has been widely used over the years (23, 25). Although this technique has been particularly useful in identifying both GAS and GBS in the clinical microbiology laboratory, its utilization has some drawbacks which negatively influence the classification of GCGS. It is now recognized that GCGS strains related at the species or subspecies level may have different Lancefield antigens. The most notorious example is found in SDE for which isolates bearing Lancefield group G, C, L and A antigens are known (32, 33). Moreover, distinct species may bear the Lancefield group C or G antigens (25), including certain members of the unrelated "*S. milleri*" group, which are distinguishable from GCGS by the formation of minute colonies after growth on solid media for 24 hours at 35 to 37ºC (22) (Fig. 1.1).

Contemporaneously with the Lancefield grouping scheme development, it was noted that the serologically defined groups could also be divided by means of physiological

tests (23). The attempted sub-grouping of GCGS into individual species was subsequently based on their serological properties, biochemical testing results and host distribution (3). The GCGS classification obtained by combining these data changed over the years and distinct epithets were used to name the defined groups. More recently, the application of genotypic methods allowed the taxonomy of the group to be clarified (25).

1.2. Species delineation and host distribution

GCGS is a phenotypic and genotypically diverse group comprising several taxonomic entities which have evolved in different hosts. In some situations, this has lead to a niche separation which correlates with the taxonomic divisions accepted at the present time. Seven distinct *taxa*, including four species, are currently defined (Table 1.1). Further division into subspecies resulted from the genotypic analysis which justified their inclusion under the same epithet (27, 34). GCGS speciation is important because the individual *taxa* differ in their pathogenic characteristics, may cause distinct infections in the same host, present different infection transmission routes and their prevalence in human infection differs markedly.

Table 1.1 shows the current taxonomic status of GCGS and summarizes key aspects such as host distribution, ecologic behaviour and spectrum of infections caused by the individual *taxa*. While some GCGS behave as host restricted pathogens, others can be regarded as generalists having the ability to infect a wide range of distinct animal species, including humans. The occurrence in human infection is well established for SDE, *S. equi* subsp. *zooepidemicus* and *S. canis* (25), though their prevalence is quite different. These three *taxa* display a broad host range and can be found as either colonizers and/or pathogens of many domestic and wild animals (33–35). SDE is by far the most frequent in human infection (17–19) and the only one for which human colonization is recognized (25). Another contrasting feature of SDE is that isolates recovered from human and non-human sources can be distinguished by both phenotypic and genotypic tests (27, 33), a separation not observed for the other two species. *S. equi* subsp. *zooepidemicus* and *S. canis* are mainly animal pathogens but their occurrence in human infection is well documented (5, 36). For the other GCGS listed in Table 1.1, infection of the human host is negligible, as very few confirmed cases of human infection have been published (37–40), the description of poorly characterized isolates

TABLE 1.1. Main characteristics of GCGS species

Species[a]	**Lancefield group**	**Host**[b]	**Recognized role as Colonizer/Pathogen**[c]	**Commonly reported infections in human/animal host**[d]
S. dysgalactiae subsp. *equisimilis*	A, C, G, L	Human	Both	SSTI, bacteremia, pharyngitis, STSS
		Animal (various)	Both	Pneumonia, septicemia, arthritis, uterine infections
S. dysgalactiae subsp. *dysgalactiae*	C	Animal (cow, fish)	Pathogen	Bovine mastitis, fish septicemia
S. canis	G	Animal (various)	Both	Dermatitis, otitis externa, pneumonia, infective endocarditis, bacteremia
		Human	Pathogen	Bacteremia, SSTI
S. equi subsp. *zooepidemicus*	C	Animal (various)	Both	Septicemia, pneumonia, arthritis, uterine infections
		Human	Pathogen	Bacteremia, meningitis, nephritis, arthritis, STSS
S. equi subsp. *equi*	C	Animal (horse)	Pathogen	Equine strangles
S. equi subsp. *ruminatorum*	C	Animal (various)	Pathogen	Ovine/caprine mastitis, strangle-like disease in hyenas
		Human	Pathogen	Bacteremia
S. phocae	C, F, G, none	Animal (seals, fish)	Pathogen	Systemic infections

[a] The *taxon S. dysgalactiae* subsp. *dysgalactiae* is not beta-hemolytic (presents gamma or alpha hemolysis). The inclusion in GCGS is justified by the genetic proximity with other species of the group.

[b] Isolates from a given species are divided in lanes human/animal to differentiate their role in colonization and infection according to host. For species other than SDE, this separation does not imply that they actually represent distinct populations.

[c] Refers to the main role. Classification as "pathogen" indicates that role as colonizer or reservoir has not been clearly identified or is restricted to a specific ecological niche.

[d] SSTI – skin and soft tissue infections; STSS – streptococcal toxic shock syndrome.

questions their true identity or, as is the case of *Streptococcus phocae*, have not yet been reported.

1.2.1. *Streptococcus dysgalactiae* subsp. *equisimilis* (SDE)

The GCGS strains currently included in the *S. dysgalactiae taxon* were long divided into groups according to their host distribution and phenotypic characteristics (27, 33). At least five groups were recognized over the years, which have been reclassified under the *S. dysgalactiae* epithet following the genetic evidence for the inclusion in a single species of the non-hemolytic group C *S. dysgalactiae* strains of bovine origin and the beta-hemolytic strains of Lancefield groups C (*S. equisimilis*), G and L (34) (Table 1.2).

TABLE 1.2. Subspecies division of *S. dysgalactiae* strains

Former *S. dysgalactiae* groups[a]	**Subspecies proposal**[b]	
	Vandamme *et al.*, 1996 (27)	**Vieira *et al.*, 1998** (33)[c]
Bovine non-hemolytic group C (*S. dysgalactiae*)	*S. dysgalactiae* subsp. *dysgalactiae*	*S. dysgalactiae* subsp. *dysgalactiae*
Animal beta-hemolytic group L strains		
Animal beta-hemolytic group C (*S. equisimilis*)		*S. dysgalactiae* subsp. *equisimilis*
Human beta-hemolytic group C (*S. equisimilis*)	*S. dysgalactiae* subsp. *equisimilis*	
Human beta-hemolytic group G		

[a] Both terms *S. dysgalactiae* and *S. equisimilis* were used in the past. Group L strains have been more frequently isolated from animals than from humans and cannot be distinguished from each other, thus being assigned as animal strains.

[b] Dashed lines indicate the separation point according with the two distinct subspecies proposals.

[c] The subspecies division proposed by Vieira *et al.* is the most commonly accepted and is the one adopted in this thesis.

The two subspecies currently defined for *S. dysgalactiae* were initially established in 1996 by Vandamme *et al.* (27). Based on whole-cell protein analysis and physiological tests, these authors proposed that all strains isolated from humans should constitute the subspecies *equisimilis*, while *S. dysgalactiae* subsp. *dysgalactiae* would accommodate all strains of animal origin, regardless of the hemolytic activity displayed. However, a study conducted two years later by Vieira *et al.* (33), based on DNA-DNA hybridization and multilocus enzyme electrophoresis data, showed that the alpha-hemolytic group C strains of bovine origin were a group distinct of all the others and restricted the *S. dysgalactiae* subsp. *dysgalactiae taxon* to these strains. Thus, these authors proposed SDE to include all the beta-hemolytic strains, from both humans and animals. This

segregation by hemolytic reaction but not by host became the most accepted (25). However, the disagreement in the two original studies on the taxonomic position of the animal beta-hemolytic strains and some inconsistencies observed in the 16S rRNA analysis of those strains (41), has led different authors to classify them in distinct manner and at the present time this classification is still a matter of debate (42).

In view of the current taxonomic status, SDE still encompasses a phenotypically diverse set of strains, which may be differentiated by their Lancefield group carbohydrates, biochemical characteristics and by host specificity. Most SDE isolates recovered from human specimens present the Lancefield group G antigen, followed by group C, while Lancefield's groups L and A are seldom found (17, 43). The occurrence of this last group carbohydrate in SDE strains is of particular significance given their potential erroneous identification as GAS. The first SDE strain bearing the Lancefield group A antigen was identified in 1997 (32), and group A SDE have been isolated and characterized in a few other studies (43, 44). A study from Japan pointed to the dissemination of a single clone in the country (44), a likely indication that this antigen may be restricted to a single or only a few SDE genetic lineages.

SDE is also found as a colonizer and pathogen in a wide variety of animal hosts, including house pets (such as dogs and cats), domesticated cattle (pigs, sheep, cows, horses, chicken) and as an opportunistic pathogen in many other wild species (27, 41, 45). SDE strains isolated from animals are predominantly of Lancefield groups C and L. While group C strains were shown to possess unique phenotypic and genotypic characteristics, allowing their separation from human SDE isolates expressing this Lancefield group, group L strains have been reported to be indistinguishable independently of the host they have been isolated from (33, 46).

The other species of the *S. dysgalactiae taxon*, *S. dysgalactiae* subsp. *dysgalactiae*, is the only non-hemolytic species included among GCGS (25). This *taxon* now includes alpha- or gamma-hemolytic Lancefield group C strains commonly associated with bovine mastitis and also isolated from the vagina of cows (33), a similar description to the one made by Garvie *et al.* in 1983 for *S. dysgalactiae* (47). Over the years, the term *S. dysgalactiae* was occasionally used to designate Lancefield group C strains isolated from animals other than cows (47, 48) and some authors described cases of human infection by *S. dysgalactiae* subsp. *dysgalactiae* (49, 50). However, such reports do not give an accurate description of the microbiological characteristics of the isolates which

could unambiguously identify them as *S. dysgalactiae* subsp. *dysgalactiae*, especially when the taxonomic issues described above are considered. At present there is no data that consistently supports a role for this species as an agent of infection in humans or other animals, and it is arguable that it might behave as a host restricted pathogen of cows.

In recent years, non-hemolytic Lancefield group C *S. dysgalactiae* strains were identified infecting several species of cultured fish in Japan and other Asian countries (51, 52), and a case of a woman who developed cellulitis and bacteremia after being stung while cleaning raw fish has been reported from Singapure (38). As suggested by the absence of hemolytic activity and the analysis of the manganese-dependent superoxide dismutase (*sodA*) gene (53), these isolates seem to belong to the *S. dysgalactiae* subsp. *dysgalactiae taxon*. However, they present unique genotypic features that distinguish them from the *S. dysgalactiae* subsp. *dysgalactiae* type strain isolated from bovine mastitis and other *S. dysgalactiae* strains isolated from mammals (51). More recently, beta-hemolytic isolates were detected in farmed fish from Brazil (54), making the exact phylogenetic position of these strains still unclear at the moment.

1.2.2. *Streptococcus canis*

The term *Streptococcus canis* was used in veterinary clinical microbiology long before the formal establishment of the species by Devriese *et al.* in 1986 (35). According to these authors, *S. canis* should include the Lancefield group G strains found causing bovine mastitis and distinct infections in dogs which were biochemically distinguishable from human Lancefield group G streptococci (SDE) (35). *S. canis* is not only part of the microbiota of cats and dogs (55, 56) but also an important pathogen for these two species and other domestic and wild animals (35, 57). *S. canis* causes from relatively mild non-invasive infections, such as dermatitis and otitis externa, to severe invasive infections (58–60). *S. canis* may be transmitted between different animal species living in proximity, as shown by its involvement in outbreaks of clinical and subclinical mastitis with bacterial shedding in milk in cattle herds (61) and in pets living in shelters (59).

The first confirmed report of human infection by *S. canis* was published in 1997 and described a case of septicemia in a 77-year-old man (62). An increasing number of reports has subsequently identified this species mainly from cases of bacteremia (17, 21,

63) and skin and soft tissue infections (36, 64). The presence of *S. canis* in human infection is thought to have a zoonotic origin, and in most of the reported cases direct contact with colonized or infected animals was proposed (62, 64). Many studies failed to detect *S. canis* among GCGS causing human infections (11, 12, 65–67), while others found it at low frequency (17, 68). A single study reported increased incidence of this pathogen among GCGS causing both invasive and non-invasive infections in a hospital in France (36), raising the possibility of increased importance of this pathogen in certain geographic regions. In contrast to SDE, little is known of the relationships between *S. canis* strains isolated from distinct hosts and of the characteristics of this population. The publication of the first *S. canis* whole-genome sequence showed the close evolutionary relationship to both SDE and GAS, bacteria which share with *S. canis* many of the genetic determinants that are thought to contribute to the colonization of distinct niches, invasion of specific tissues and evasion of the host immune system (14, 26).

1.2.3. Other GCGS species

Lancefield group C strains found in human infection that do not belong to SDE are part of the *Streptococcus equi taxon*. The name *S. equi* was first used in the late 19th century (69), to refer to diplococci causing respiratory infection in horses. Currently, it includes three distinct subspecies which are found as animal pathogens: *S. equi* subsp. *zooepidemicus* (34), *S. equi* subsp. *equi* and the recently described *S. equi* subsp. *ruminatorum* (70). While *S. equi* subsp. *equi* is a strict horse pathogen the other two infect a broader range of hosts which also includes humans.

S. equi subsp. *zooepidemicus* is a equine commensal which may opportunistically cause infections, mainly in the respiratory tract (71). Infections caused by this bacterial species in other animal species, both domestic and wild (72–74), are often severe and include meningitis (74) or hemorrhagic pneumonia (75). In the human host, *S. equi* subsp. *zooepidemicus* has been isolated from bacteremia and endocarditis (5, 76), meningitis (77), streptococcal toxic shock syndrome (STSS) (78, 79) and nephritis outbreaks (80, 81). This is the second GCGS species for which more cases of human infection have been reported (5, 77, 81).

S. equi subsp. *equi* is long known to be the etiological agent of equine strangles, a highly contagious disease that affects the upper respiratory tract of horses and is

characterized by the formation of abscesses in the lymph nodes of the head and neck (82). Although some horses can become carriers of the bacteria in their guttural pouch after the acute phase of the disease, the colonization seems to be restricted to this anatomical region and it is not associated with infections other than strangles. Another contrasting feature to *S. equi* subsp. *zooepidemicus* is the host range since it has not been found in other animals. Thus, *S. equi* subsp. *equi* is regarded as a highly adapted pathogen of the equine respiratory tract and it is not considered an agent of human infection (25), from which it has been very rarely isolated (4, 37). Genetic analysis of *S. equi* subsp. *equi* isolates suggests that this bacterium is a clone which evolved from *S. equi* subsp. *zooepidemicus* to become a horse specific pathogen (83).

The third subspecies of *S. equi*, *S. equi* subsp. *ruminatorum,* was only described in 2004, from ovine and caprine mastitis in Spain (70), and subsequently recognized in hyenas, zebras and wild dogs in Africa (84, 85). Soon after its description, two independent cases of invasive human infection were published (39, 86). The reservoir for this bacterium is still unknown since the niches it may occupy as a colonizer were not identified. Although the literature is still scarce for *S. equi* subsp. *ruminatorum,* the apparent distribution by host species seems to indicate that this species may be more similar to *S. equi* subsp. *zooepidemicus* than to the restricted host tropism presented by *S. equi* subsp. *equi*.

A very limited number of studies have focused on *S. phocae***,** a bacterium first described by Skaar and colleagues in 1994, as an opportunistic pathogen of seals with a viral infection (87). *S. phocae* has been later detected infecting several marine mammals (88, 89) and aquaculture salmon in Chile (90). The original study identified strains bearing antigens of Lancefield groups C, F or non-groupable, but subsequent studies reported the occurrence of the Lancefield group G antigen in this species (89, 91). No case of human infection by *S. phocae* has been reported so far. From the currently available literature on *S. phocae* it should be noted that although *S. phocae* isolates present biochemical traits that should allow their distinction from other GCGS species (87), the absence of biochemical profiles specific for this species in databases of currently used identification systems hamper its correct identification. The application of 16S rRNA gene sequencing is the most frequently used method for this species identification (89–91). Thus, infection by this agent may pass unrecognized unless molecular characterization is used.

1.3. Laboratory identification

The most commonly reported procedure for identification of a beta-hemolytic streptococcal isolate as belonging to the GCGS group involves Lancefield group determination, in laboratories performing this serological technique, or identification by means of biochemical testing. However, according to the available literature, in cases of beta-hemolytic pathogens other than GAS and GBS, additional characterization of GCGS isolates, including species level identification and antimicrobial susceptibility testing, may be relatively uncommon (17). The serological identification of these Lancefield groups will imply in first instance the confirmation of a large-colony-forming phenotype. In addition to colony size, the Voges-Proskauer test (acetoin production) will allow the differentiation of GCGS from the small-colony-forming "*S. milleri*" strains bearing identical Lancefield groups (25), as only the members of the latter give a positive reaction in this test.

Most of the GCGS *taxa* can be distinguished by biochemical testing, including the identification of the subspecies defined for *S. dysgalactiae* and *S. equi* (82). Among the carbohydrate fermentation tests allowing species identification, the fermentation of trehalose and sorbitol (22) identify Lancefield group C strains as SDE, *S. equi* subsp. *zooepidemicus* or *S. equi* subsp. *equi* (25). On the other hand, the isolation of Lancefield group G streptococci from a human specimen requires the distinction between SDE and *S. canis*, which can be achieved by the positive reactions of beta-glucoronidase activity and acid production from trehalose for the former, and alfa- and beta-galactosidase activity for the latter (21). SDE isolates which bear the Lancefield group A antigen can be distinguished from GAS by testing pyrrolidonyl arylamidase (or aminopeptidase) activity, since only GAS produces this enzyme (19, 21). Some commercial kits have also proved to be useful in identifying GCGS species, namely SDE (65). However, the phenotypic tests described above not always allow the accurate identification of all *taxa* currently included in the group.

Other methods have also been applied, including matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS), which is being increasedly used in the clinical microbiology laboratory as it provides fast, accurate and cost-effective bacterial identification (92). Few studies have evaluated the capacity of this technique, based on the protein composition of microbial cells, to specifically categorize individual GCGS species. Although one report applying this technique to

beta-hemolytic streptococci showed that species level identification of *S. dysgalactiae* may be accomplished (92), another one, which used 16S rRNA gene sequencing as the gold standard, reported the misidentification of a SDE isolate (93). Thus, genotypic methods may be the only way to warrant exact identification in many instances.

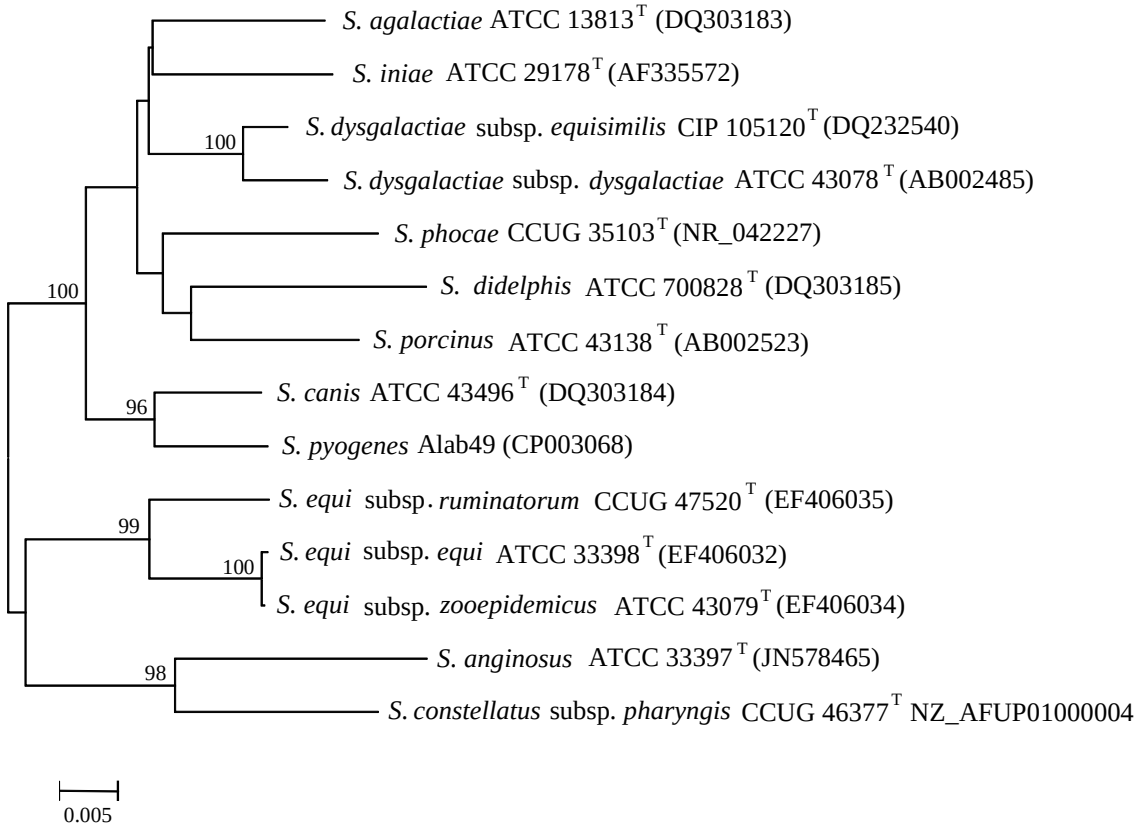

FIGURE 1.2. Neighbor-joining tree of the 16S rRNA gene of GCGS and other beta-hemolytic species

Branches clustering sequences with greater that 75% bootstrap support (1,000 replicates) are indicated. The tree is drawn to scale, with branch lengths corresponding to the number of base substitutions per site. The species name and strain identification are indicated. GenBank entries whose sequences were used to construct the tree are shown in brackets. Small-colony-forming "*S. milleri*" species which may also possess Lancefield group C or G antigens are also included. "^{T}" denote type strain.

Several molecular approaches have been employed for the differentiation of GCGS, but only a few were aimed at the routine laboratory identification, while others were directed towards veterinary microbiology (94). A few studies had a special focus on assays allowing the distinction between SDE and GAS (95, 96). Among the sequence based methods developed for streptococci and applied in GCGS speciation, several distinct genes were targeted (21, 41, 42, 53). Among these, sequence analysis of the 16S rRNA gene should be highlighted, because it was important in defining the current

taxonomic status of the genus (25) since it allows the distinction of all the currently defined GCGS species (Fig. 1.2), and it has been the most used method by workers in the field (65).

Disagreements were frequently seen among phylogenies derived from distinct genes (41, 42, 97). For example, 16S rRNA gene analysis places *S. canis* as the closest relative of GAS (as observed in Fig. 1.2), while analysis based on other housekeeping genes shows a closer relationship between SDE and GAS (97). Multilocus analysis and full genome sequence data have confirmed a close relationship for the three *taxa* (26, 42), but SDE was identified as most likely sister species of GAS (26).

2. GCGS ECOLOGY AND EPIDEMIOLOGY IN THE HUMAN HOST

The recognition of SDE as part of the human microbiota and the diversity of clinical conditions for which this species is believed to be responsible, makes SDE the most relevant GCGS species in the context of the human host. Although some older literature has limited information on GCGS species identification, more recent studies focusing on the clinical and epidemiologic features of SDE infections produced similar conclusions, confirming previous observations and supporting a major role of this species in the earlier studies.

2.1. Ecological niches and origin of infection

The occurrence of GCGS in human specimens is known since the original description of Lancefield groups C and G bearing strains by Rebecca Lancefield (2). Sites associated with colonization by these organisms include the nasopharynx, the skin, the gastrointestinal and the female genitourinary tracts (2, 98, 99). Although there is a limited number of reports on GCGS carriage in the human host and many did not identify the species, it is generally accepted that SDE is the only GCGS that can consistently colonize humans (100).

Most of the available studies on human GCGS colonization have focused on the pharynx. SDE asymptomatic carriage rates of 2 to 3% have been reported in both children and young adults (98, 101–103). Throat colonization rates have been noted to change over time (101) and others have shown they may be much higher in certain human populations (104–106). A study with school children from India found a SDE colonization rate of almost 10% (106), and in the aboriginal communities of Northern Australia this value may go up to 20% (104, 105).

The origin of the infection has been poorly studied in SDE. Taking into consideration the presence in the human microbiota and the potential separation of human and animal strains within this *taxon* (33), the most likely source of infection is endogenous. This idea is reinforced by the fact that SDE infections often occur in the skin and upper respiratory tract (19), suggesting that colonization of these areas may have a role in the origin of infection. For example, the presence of these bacteria in toe webs (107) or anal colonization (99) was suggested as reservoirs for erysipelas and cellulitis of the lower limbs. SDE is likely present in the intact skin before the onset of infection and

cutaneous infections, mainly those involving deeper layers of the skin (e.g. cellulitis), may allow SDE to access the bloodstream and cause invasive infections (7, 108). Indeed, studies characterizing SDE bacteremia have repeatedly implied the skin as the primary focus (following not only cellulitis, but also wounds, ulcerations and intravenous drug use) (4, 12, 17, 109). The origin of infection may also be localized to the upper respiratory tract (in neutropenic patients and those with pneumonia), and the genitourinary tract (in women during the peripartum period) (4, 68).

A person-to-person route of transmission has also been demonstrated for SDE. In a case-control study conducted in Finland (67), SDE was more frequently isolated from household members of patients with cellulitis. Molecular typing identified the same SDE strain in the pharynx of patients and members of their household. References can also be found in older literature to healthcare-associated infections associated with both person-to-person transmission and transmission from contaminated environmental sources, with SDE being implicated in outbreaks of skin infections and puerperal sepsis (110) and in post-operative wound infections (111). Less frequently, food borne community outbreaks of pharyngitis attributable to SDE have been detected (110), although this route of transmission is more common for *S. equi* subsp. *zooepidemicus* (see below).

The phenotypic and genotypic separation between SDE strains isolated from humans and animals (46) could be interpreted to suggest that zoonotic transmission is not frequent in this *taxon*. However, no separation between group L strains isolated from human and non-human sources is apparent (33) and a recent report identified the same SDE strain in a child and a dog (112). These observations show that cross-species transmission can occur and question whether animal SDE strains are transmitted to humans, as observed for other GCGS species. It should be noted that infections by such strains may pass essentially unrecognized as neither the genotypic characteristics of such isolates are firmly established, nor is the intra-specific characterization of SDE isolates that could allow such discrimination to be made usually performed.

The other GCGS species are not recognized as agents of colonization in humans. Nevertheless, *S. equi* subsp. *zooepidemicus* has been identified in throat swabs of farm workers with no clinical signs of infection at farms experiencing outbreaks caused by this pathogen, and also in convalescent patients several months after the original infection (80). No such evidence have been found for any other of the GCGS species

that are present in animals, but one study detected an unusual high number of *S. canis* isolates in non-invasive human specimens suggested the occurrence of human colonization by this agent (36). Thus, it is conceivable that transient colonization by *S. equi* subsp. *zooepidemicus* and *S. canis* may occur in people contacting frequently with animals.

Human infection by GCGS species other than SDE, as is the case of *S. canis*, *S. equi* subsp. *zooepidemicus,* and more recently, *S. equi* subsp. *ruminatorum*, is thought to originate from contact with animals or with their products. For these three species, cases of direct transmission through pre-existing skin lesions, such as patients' wounds and ulcers, or direct inoculation by bites or scratches, were considered the most likely portal of entry (63, 113, 64, 114, 86). A couple of studies employed typing methods to definitively determine the zoonotic origin of the infection in these patients. One report identified the same *S. canis* strain in a dog and its owner's blood cultures following a bite (63), and another one showed the transmission of a *S. equi* subsp. *zooepidemicus* strain from a dog to a handler (115). Only *S. equi* subsp. *zooepidemicus* infection has been associated with the consumption of unpasteurized milk products, resulting in outbreaks of bacteremia (116) and glomerulonephritis (81). The presence of this bacterium in milk is related to its role as an agent of mastitis in cows and other cattle (80, 117). Although other GCGS species may cause mastitis in animals with bacterial shedding in milk (47, 61, 70), such a route of infection has not been established.

2.2. Infections

GCGS have been associated with a multitude of infections in humans, ranging from mild to life-threatening conditions. Most commonly, infections occur in the skin and in the upper respiratory tract (3, 19) The spectrum of infections caused by this species closely resembles that caused by GAS, and includes numerous skin and soft tissue infections (67), pharyngitis (98), ocular infections (118), bacteremia (119), endocarditis (120), septic arthritis (121), pneumonia (122), meningitis (110), necrotizing fasciitis (123) and streptococcal toxic shock syndrome (STSS) (124). *S. canis* has been mostly isolated from skin and soft tissue infections and associated bacteremia (21, 36), while *S. equi* subsp. *zooepidemicus* infections are often severe, including septicemia (125), endocarditis (76), septic arthritis (126) and meningitis (77), sometimes with the

formation of brain abscesses (113). This bacterium has also been associated with STSS (78, 79) and with several outbreaks of acute glomerulonephritis (80, 81).

Skin and soft tissue infections are usually the most common SDE clinical manifestation and may range from wound infections (111), impetigo (104), erysipelas (127), and cellulitis (67), to severe conditions such as necrotizing fasciitis (123). Milder forms of skin and soft tissue infections and cellulitis occur predominantly in the lower limbs (7), and disruption of the cutaneous barrier may serve as site of entry. Recurrence has been noted for cellulitis (67).

SDE has been frequently identified from throat swabs (106, 128, 129), but its role as an etiological agent of pharyngitis has been questioned. Factors that may explain this include the fact that SDE may be present in a significant proportion in the throat of healthy people (98), identification issues related to the differentiation from "*S. milleri*" strains (130) and potential differences in importance in patients of distinct age groups (102). Studies conducted in young adults in the United States showed that patients from whom SDE was isolated in throat swabs had clinical features compatible with pyogenic infection (129) and a statistic association between SDE and the presence of symptoms compatible with pharyngitis in a similar population has been reported (131). However, other studies failed to detect such association in the pediatric population (102, 103, 106). A study from Norway concluded that although patients with GAS tonsillitis are younger, patients with GCGS tonsillitis present the same clinical picture and that SDE should be considered as throat pathogen (132). Pharyngitis epidemics caused by SDE have long been recognized and to affect people of all age groups (110, 133). Outbreaks due to *S. equi* subsp. *zooepidemicus* after consumption of contaminated milk products have also been recognized (80). In the case of the latter species, these are normally followed by more severe clinical manifestations (5, 81).

Most of the published literature on GCGS infections report invasive infections (4, 11, 65, 108, 120, 124). In the last ten years, numerous studies described SDE bacteremia episodes, providing some form of epidemiologic characterization of the isolates and clinical features of the patients. The most common clinical manifestations associated with SDE bacteremia included primary bacteremia, septicemia, cellulitis, endocarditis, septic arthritis, osteomyelitis, pneumonia, meningitis, necrotizing fasciitis and STSS (12, 17, 68, 122). Disruption of the cutaneous barrier is common in patients with SDE bacteremia (11, 12) and bacteremia secondary to cellulitis is a common finding (7, 65,

108). Those studies also identified many underlying diseases among patients and noted a frequent recurrence of bacteremia (7, 32, 68, 119). *S. equi* subsp. *zooepidemicus* has also been identified from bacteremia (5) and endocarditis (76), while *S. canis* has been isolated mostly from sporadic bacteremic episodes (62, 63).

The first description of STSS caused by a Lancefield group G strain was in 1996 (9), and since then, many other reports have confirmed SDE involvement in this severe syndrome (17, 78, 124, 134, 135), for which mortality rates of nearly 50% have been reported (136). As noted for SDE bacteremia, STSS patients commonly have underlying diseases (136). *S. equi* subsp. *zooepidemicus* has also been identified in cases of human STSS (78, 79). *S. canis* has not yet been associated with either necrotizing fasciitis or STSS in humans, although it does cause infections with similar features in house pets (58).

GCGS have also been associated with the occurrence of non-suppurative syndromes typically associated with GAS. Definitive epidemiologic evidence for *S. equi* subsp. *zooepidemicus* in triggering glomerulonephritis has been obtained. Duca *et al.* (80) described the first outbreak of sore throat followed by acute glomerulonephritis caused by this species in Romania in 1969, due to the consumption of contaminated milk from cows with mastitis (80), and subsequent studies implied this organism in glomerulonephritis outbreaks, always associated with consumption of raw milk products (81). The hypothesis that Lancefield group G streptococci could also trigger this disease has been proposed almost thirty years ago (101), but very few cases have been reported (110).

GCGS were also proposed to trigger acute rheumatic fever and rheumatic heart disease in certain human populations experiencing simultaneously high incidence of these diseases and high rates of GCGS pharyngeal carriage (105, 106). These diseases, believed to result from an autoimmune response subsequent to an upper respiratory tract infection by GAS (137), are very frequent in the Aboriginal population from northern Australia, even though GAS is rarely isolated from either carriage or symptomatic infection at the pharynx (105). Hadean *et al.* showed that GCGS could elicit a immune response leading to acute rheumatic fever, as antibodies against GCGS isolates showed cross-reactivity with human heart myosin, which has been identified as a major cardiac antigen recognized by heart cross-reactive autoantibodies against GAS (105). Although later studies confirmed important throat carriages of SDE in these communities (138),

and others explored the possible role of the SDE M-like protein in triggering this syndrome (139), no definitive evidence has been generated so far for the role of SDE or other GCGS species in triggering these diseases.

2.3. Clinical and epidemiologic features of GCGS infections

SDE is clearly the GCGS species responsible for most cases of both invasive and non-invasive infections in the human host reported worldwide (17, 140–142). SDE has been frequently the only species found among Lancefield group G isolates causing human infection (12, 65), while *S. canis* generally accounts for at most only a small proportion of the isolates (17, 68). Similarly, SDE is the most commonly isolated species among Lancefield group C isolates (17, 19, 120), although some studies also report the presence of *S. equi* subsp. *zooepidemicus* (4, 5). A study from Hong Kong showed that this species was the most prevalent among Lancefield group C isolates causing bacteremia (5). In a ten-year study in Finland (12), the only *S. equi* subsp. *zooepidemicus* isolated could be related to an outbreak associated with consumption of unpasteurized cheese (116), showing that increased detection of this species in some of these studies may be due to specific epidemiologic situations.

The disease burden attributable to GCGS is usually inferior to that caused by GAS and GBS in most geographic locations (7, 143), but SDE appears as the most prevalent beta-hemolytic pathogen after those two species (17). However, it has been noted that colonization rates and the number of cases of both non-invasive and invasive infections may occasionally exceed that of GAS and/or GBS. Throat colonization rates (104–106) and isolation from patients with clinical signs of pharyngitis (106, 144) exceeding that of GAS, were reported from India, Australia and the United States. In Finland, SDE was shown to be the main beta-hemolytic species causing erysipelas and cellulitis in hospitalized adults (67) and bacteremia in adult patients (12).

For invasive infections, population-based studies have shown GCGS annual incidence rates of 2.24 per 100 000 population in Canada (145), a similar value in United States (17), and up to 4.3 per 100 000 in Finland (12). Mortality rates ranging from 15% (17, 65, 109) to 25% (4) have been reported for GCGS invasive infection, although in patients with STSS this value may be higher. Most of the published series report community acquired infection (65, 108, 122), but up to one third of the bacteremias may be related to hospital care (109). The incidence of GCGS bacteremia

was identical to that of GAS but lower than that of GBS in the United States (120), and has occasionally been reported to exceed that of GAS in Israel (11) or that of GBS in adult patients from Denmark (109). In Japan, where numerous reports of invasive infection by SDE have originated, including STSS (6, 78, 136), invasive infections by this species have increased gradually (6, 136) and SDE is now responsible for a disease burden superior to that of other beta-hemolytic species (122). Among the studies showing relevant incidence of SDE bacteremia, several pointed to an increasing trend over time. Such patterns were noted over the 1990's in one hospital from Israel (11) or over a 10-year period in the Finish study cited above (12), for example.

Despite the similarity in the spectrum of infections caused by SDE and GAS, differences in severity and in the clinical characteristics of the patients have been noted by some authors. Patients with SDE invasive infections were reported to be older and to have more co-morbidities than GAS patients (109, 122). Most of the patients with SDE invasive infections have underlying diseases, including malignancy, diabetes mellitus, cardiovascular disease, peripheral vascular disease, immunosupression, substance abuse, chronic renal failure, chronic wounds and immobilization (4, 11, 17, 65, 68, 119, 146). Given that most patients are elderly and often present with multiple underlying conditions (4), immunosenescence has been proposed as the major risk factor for SDE invasive infection (108). Although this view highlights the opportunistic nature of SDE infections, this species has been shown to have the ability to cause severe invasive disease in previously healthy adults (123). The multiple examples of increased isolation rates of GCGS described above could reflect improved detection methods or an increased virulence of the bacteria. Given the importance of underlying conditions for the infection by these microorganisms, an expanding population of compromised hosts, more susceptible to infections by these agents, should also not be disregarded as a cause for their increasing importance in human infection.

2.4. Virulence factors

The pathogenic role of GCGS for their respective hosts is highlighted by the multitude of factors now known to be encoded in their genomes. These allow GCGS to specifically bind or degrade host components, enabling them to adhere and spread through cellular tissues and to evade the host immune system. A large body of information on the virulence factors carried by SDE has been provided by studies trying

to identify such factors in GCGS isolated from human infections, while studies focusing on the pathogenic properties of GCGS species of veterinary relevance also accomplished a good characterization for members of the *S. equi* group (71).

The ability of GCGS to exert lytic activity on human fibrin was recognized shortly after the identification of GCGS (23), and subsequent studies showed that these bacteria could bind many human components (147, 148). Cloning and sequencing of the genes encoding the proteins responsible for these properties led to the identification of several potential virulence factors, as was the case of protein G, a surface molecule with immunoglobulin-binding properties (149). Another common approach was to look for factors linked previously to virulence in GAS, given the recognized similarity between the two species. Several homologs of GAS genes were identified in SDE, including the gene coding for the M protein (150), one of the most important virulence factors in GAS (137). More recently, microarray and genome analysis confirmed that many of the virulence factors present in both species are shared (15, 151). Whereas the extensive study of the mechanism of action of these factors was mostly undertaken in GAS, recent studies in SDE and other GCGS species confirmed their contribution for the pathogenesis of these microorganisms and, in some instances, revealed unique features of the GCGS counterparts. Differences in the presence of particular *loci*, in the allelic sequences of universally distributed genes or differential gene regulation, can all be assumed as mechanisms justifying the differences in virulence observed between GCGS and GAS.

Two types of virulence factors have deserved special attention over the years, following their recognized importance in GAS pathogenesis: M protein and superantigens (streptococcal pyrogenic exotoxins). In the case of the M protein, not only is a similar protein encoded in SDE genomes, but a number of M-like proteins (*i.e.* proteins presenting similar structure but that may be encoded in distinct *loci*)[3] have been detected in other GCGS species (152, 153). Functional characterization conducted in those GCGS species pointed for the importance of M and M-like proteins in their virulence in distinct hosts. In contrast, limited evidence supports the contribution of

[3] The term "M-like proteins" refers to proteins presenting a similar structure to the archetypal M protein encoded by the gene *emm* in GAS. In GAS, a number of M-like proteins are encoded in distinct *loci* which are referred to as *emm*-like genes. Similar proteins are also found erratically distributed in GCGS species and are often referred to as M-like proteins, although the exact genomic localization of these genes has not been determined in many instances.

superantigens for SDE virulence in the human host, despite the high number of studies looking for GAS genes in this species.

2.4.1. M and M-like proteins

The GAS M protein is the archetypal of a group of structurally similar cell-associated proteins with antiphagocytic or immunoglobulin binding properties (154). Its fibrillar coiled-coil structure is thought to be essential for the multifunctional character which is known to effectively contribute to invasiveness by mediating resistance to phagocytosis, cell adherence and internalization of GAS strains (137, 154). Following the observations that these bacteria resisted to phagocytosis *in vitro*, reacted with serum antibodies raised against the GAS M protein and had fimbriae like structures at the cell surface, as shown by electronic microscopy (155, 156), the presence of a protein similar to the GAS M protein was confirmed by biochemical and genetic studies (150, 157).

The M protein is encoded by the *emm* gene, which is present in virtually all GAS and SDE isolates. Thus, the *emm* genes of these two species are regarded as orthologs (*i.e.* they evolved from a common ancestral gene) and share structural features, such as substantial polymorphisms at their 5′ end (157). This portion of the gene determines the serological specificity of the M protein and is targeted by the *emm* typing technique widely used in the epidemiologic characterization of both species (158). SDE and GAS M proteins have similar structures, both presenting conserved C-terminal and hypervariable N-terminal parts (150, 159) (Fig. 1.3a). Two major classes of M proteins, I and II, have been proposed in GAS based on the epitopes within the C-repeat region located adjacent to the carboxy-terminal side of the pepsin cleavage site (154), and the SDE M protein was shown to have high sequence similarity to GAS class I M proteins in this region (150). The SDE M protein was shown to bind fibrinogen (via its N-terminal part), human serum albumin (via the C-repeats), IgG, IgA and plasminogen, and to present cross reactivity to keratin and to a joint cartilage protein (160, 161).

A few detailed studies were conducted in the last years characterizing the biologic properties of a SDE M protein (161, 166–169).[4] These studies proposed the M protein

[4] The name FOG has been proposed by the authors for the studied protein. Primers used targeted the flanking regions of the *emm* gene described in other studies and, although not clearly stated by the authors, there is no evidence that a *locus* distinct of the *emm* gene encoding the M protein in the SDE genome is being targeted.

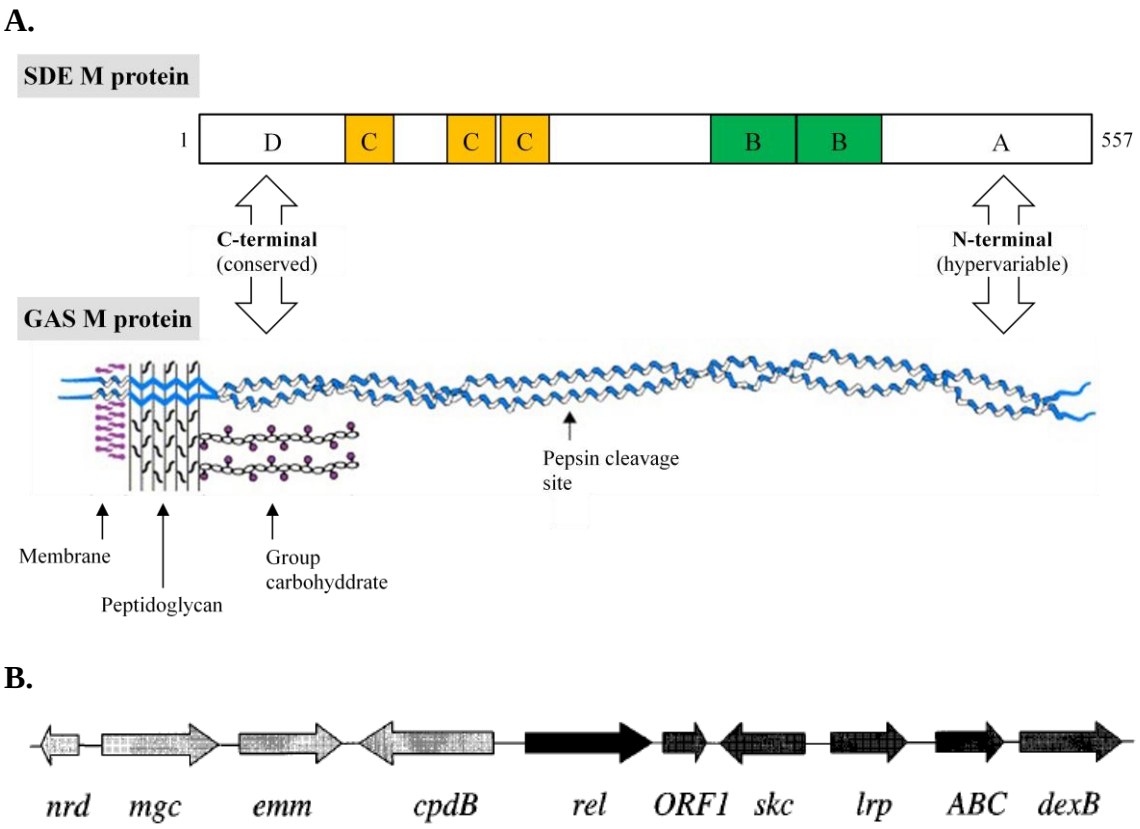

FIGURE 1.3. M protein structure and genomic environment of the *emm locus* in SDE

(A) Comparison of SDE and GAS M proteins. The depicted SDE M protein structure is the one reported for protein FOG, corresponding to a *emm* type *stG11* isolate (161), and was adapted from this reference. A to D domains are indicated in the schematic representation of this M protein. M proteins in SDE isolates containing one to five C-repeats have been described (162). Numbers refer to amino acid residue positions, where 1 is the first residue of the mature molecule. The figure of the GAS M protein was adapted from reference (163). The SDE and GAS M protein representations are shown in the corresponding relative positions.
(B) Genomic location of the *emm locus* in SDE. Two types of multigene regulator-like regions have been described for SDE, which differ by the presence or not of the *cpdB* gene (164, 165). The following genes are indicated: *nrd* – ribonucleotide reductase, *mgc* – multigene regulator-like, *emm* – M protein, *cpdB* – 2'3' cyclo-nucleotide 2'-phosphodiesterase , *rel* – protein that functions in the synthesis and degradation of guanosine 3',5' bipyrophosphate (ppGpp), *skc* – streptokinase, ABC – ABC transporter, dexB – alpha-glucosidase-like protein, *ORF1* and *lrp* – proteins of unknown function. Reproduction from reference (164).

would mediate survival and multiplication in whole blood, contribute to evasion of the immune system by forming complexes with fibrinogen causing neutrophil aggregation (161) and to promote adhesion to host tissue by binding to collagen I (168). They also showed unique features of the M protein in face of the host innate immune responses compared to protein G even if both proteins interact with the same human components (166, 169). In a distinct line of investigation, the interaction of the M protein with collagen has also been explored to infer the potential of SDE strains to trigger acute rheumatic fever due to collagen autoimmunity (139). Taken together, the currently

available data on the SDE M protein strongly suggests that it plays an important part in the pathogenesis of SDE.

The genomic context of the *emm* gene in SDE is distinct of GAS. In the latter species, the *emm* gene is located within a contiguous pathogenicity island referred as the *mga* or *vir* regulon (170), where other virulence-related genes are encoded, including *emm*-like genes (such as *mrp* and *enn*) or the gene encoding the C5a peptidase. All these genes are coordinately regulated by a positive multigene regulator encoded by the *mga* gene, located upstream of the *emm* gene (170). In SDE, an *mga*-like gene (called *mgc* in SDE) is also found upstream of the *emm* gene, presenting around 61% sequence similarity to GAS *mga* (164) (Fig. 1.3b). Whether the product of the *mgc* gene functions as a regulator in SDE is unknown. In clear contrast to GAS, neither *emm*-like genes nor the C5a peptidase gene are present in this location, but downstream of the *emm* gene in SDE is the region encoding the streptokinase, another important virulence factor (164). Recent genome analysis of GAS, SDE and also *S. dysgalactiae* subsp. *dysgalactiae*, mapped an ancient pathogenicity island including the *emm* region and extending for up to 47-kb in all GAS genomes (171), whereas in both *S. dysgalactiae* subspecies most of the genes are instead present in subsets dispersed throughout the genome (172).

A number of M-like proteins have been described from other GCGS species. Although in *S. dysgalactiae* subsp. *dysgalactiae emm*-like and *mga*-like genes are found physically associated (152), suggesting that the same *locus* is present in both GAS and SDE, no such evidence has been reported for other GCGS. However, these species do possess proteins that have functional and structural features of the M protein family. M-like proteins have been described for the three subspecies of *S. equi* (85, 153). These proteins differ from each other in terms of both sequence and function (153) and have higher similarity to M-like proteins which are encoded in *loci* other than the *emm locus* in GAS genomes. More recently, a M-like protein was also described in *S. canis* (173), which was shown to mediate the binding of human plasminogen and fibrinolytic activity (174), and full genome analysis of a representative of this species confirmed the presence of several *emm*-like genes (14).

2.4.2. Superantigens

GAS pyrogenic exotoxins act as superantigens and are thought to contribute to STSS caused by this species (163). The association of most of these genes with phages in

GAS, which could potentiate their horizontal transfer, and the implication of SDE in numerous cases of STSS, prompted the idea that such genes could also be present in the latter species. However, studies specifically looking for these genetic determinants in STSS SDE isolates were unable to detect any other than the *speG* gene that is not phage associated (78, 136).

Some SDE isolates possess an ortholog of the GAS *speG* gene in their genome (175) and this seems to be almost the only one present in SDE. Studies which looked for the *speG* gene in SDE human isolates in distinct geographic locations detected it in variable proportions. For example, *speG* was detected in 43% of the isolates causing STSS and necrotizing fasciitis in Japan (136) and in 58% of invasive and non-invasive isolates in a recent study from Israel (176). Data has been accumulating that *speG* is not expressed in SDE (177), probably due to mutations in the promoter region (178), while another superantigen chromosomal gene in GAS, *smeZ*, appears to have been lost from the SDE genome during evolution (178). One study conducted in Norway identified a non-invasive SDE isolate that carried *speC, speG, speH, speI* and *smeZ* (140). These genes presented identical or near identical sequences to their GAS counterparts, as well as to *speH* and *speI* found in *S. equi* subsp. *equi* (140), suggesting that this SDE isolate may have acquired by horizontal gene transfer those genes which are phage-associated in both GAS and *S. equi* subsp. *equi* (163, 179).

Given the currently available data, it is unlikely that superantigens play a significant role in the pathogenesis of SDE infection in humans, but the same does not hold true for GCGS isolated from animal infections. A study characterizing the allelic variants of the SpeG protein among strains of *S. dysgalactiae* indicated that these may be associated with differences in mitogenic activity and that SpeG may be relevant in animal rather than human infections (180). This view is also supported by the growing number of superantigen genes detected in GCGS species associated with animal infection, including those which have counterparts in GAS, as found in *S. dysgalactiae* subsp. *dysgalactiae* (181) and *S. equi* subsp. *equi* (179), or novel superantigen genes, as those recently described in the *S. equi* subsp. *zooepidemicus* genome (182). By opposition, no exotoxin genes were detected in the published *S. canis* genome (14), although a *S. canis* human isolate carrying a new allele of the *smeZ* gene was reported in the past (183). A bacteriophage-encoded mitogen was also detected in dog isolates of this bacterium, but

its genetic determinant was not related to any of the streptococcal pyrogenic exotoxin genes (184).

2.4.3. Other factors

The number of putative virulence factors detected in SDE has been increasing over the years. Many were recognized for a long time and include cell-surface associated proteins, such as the C5a peptidase (185), the fibronectin-binding protein Fnb (186) or the plasmin(ogen)-binding protein GapC (187), and a number of extracellular proteins, including streptokinase (188), streptolysins O and S (24) and streptodornase (189). Well known GAS regulatory systems are also present in SDE, as is the case of the two-component signal transduction systems covRS and fasCAX (190), indicating that common mechanisms could mediate the virulence of both species. Regardless of the number of virulence factors that SDE shares with GAS, a few that are believed to be critical for GAS virulence are not conserved in SDE. In addition to the differences in superantigen gene content, as discussed above, SDE lacks the gene encoding the cysteine protease SpeB, and some reports have shown that the full *hasABC* operon required for the biosynthesis of the hyaluronic capsule is not conserved among SDE strains (15).

FCT-like regions (for fibronectin-binding, collagen-binding and T antigen) have been recently identified in the SDE genome (15). These are known to encode pilus-like structures in GAS (191). The SDE FCT-like regions are similar to those previously identified in GAS, GBS and *S. equi* subsp. *zooepidemicus*, and include transcriptional regulators and a number of genes which seem structurally organized in operons (15). Together with old observation of the presence of T antigens in SDE isolates (192), these data strongly suggests that SDE expresses pilus-like structures, although their distribution in the SDE population remains unknown.

A limited number of potential virulence factors were identified in *S. canis* prior to the whole-genome characterization of an isolate belonging to this species. The presence of proteins similar to the streptolysin O and M proteins of GAS in dog isolates was suggested in the past (58), but definitive evidence for the presence of M-like proteins in *S. canis* was published only recently (173, 174). The recently available genomic data indicates that this species also shares a number of virulence factors with GAS, SDE and

other streptococcal species (14), including streptolysins O and S. A gene for CAMP factor, characteristic of GBS strains, is also found in the *S. canis* genome (14).

3. ANTIMICROBIAL SUSCEPTIBILITY AND TREATMENT

The antimicrobial susceptibility patterns of GCGS species are similar to those reported for the generality of beta-hemolytic streptococcal species. SDE and other GCGS remain consistently susceptible to penicillin and other beta-lactam agents, and penicillin is a first choice antibiotic for the treatment of infection by these bacteria (3, 20). The addition of an aminoglycoside to penicillin has long been advocated for the treatment of serious GCGS infections, to avoid delayed or poor responses due to treatment with penicillin (193). The combination of penicillin with an aminoglycoside or a cephalosporin has been frequently used in the treatment of bacteremia by GCGS (4). Macrolides and lincosamides are appropriate alternatives to penicillin in the treatment of GCGS infections (20, 194), particularly in cases of patients which are allergic to the latter agent. In Portugal, where only parenteric penicillin is available, the oral route of administration of macrolides and lincosamides is an additional advantage. As observed for beta-lactams, GCGS isolates resistant to antimicrobial agents such as the glycopeptides, linezolide and quinupristin-dalfopristin have not been recognized (195).

A few reports have described GCGS isolates displaying *in vitro* antimicrobial tolerance to both penicillin (193, 196) and vancomycin (197). This phenomenon was defined for isolates presenting low minimum inhibitory concentration but an incremented (often by more than 100 times) minimum bactericidal concentration (198). More recently, a report studying SDE from both Europe and United States described isolates with increased minimum inhibitory concentration to penicillin (up to 0.25 µg/mL) (199). Although it has been suggested that these observations could be related to treatment failures and SDE have been implicated in such cases (200, 201), the significance of these observations is unclear and seems to have limited expression.

The resistance patterns detected so far in GCGS are similar to those reported for other beta-hemolytic streptococci. Resistance to tetracyclines (202), followed by resistance to macrolides and lincosamides (203), and less frequently to aminoglycosides (204) or fluoroquinolones (199), have been described. The available studies demonstrated that most of the resistance determinants are shared with other streptococci, possible a result of the association of resistance genes to some of these antimicrobial classes with mobile genetic elements that can promote their transfer

between distinct species. Examples include macrolide (205, 206), tetracycline (207, 208) and aminoglycoside (204) resistance determinants. Specific gene variants may also be present, as observed for macrolide resistant SDE isolates (209, 210).

The use of macrolides and lincosamides in the treatment of infections caused by GCGS and other beta-hemolytic streptococci has been compromised by substantial rates of resistance found in the group (211, 212). Members of these two antimicrobial classes as well as streptogramin B, have identical mechanisms of action, by inhibiting protein synthesis by binding to the 50S ribosomal subunit (205). The overlapping binding sites on the ribosome imply that resistance to these antibiotics also arises in a similar way (205). Resistance to macrolides and lincosamides in GCGS is conferred essentially by identical mechanisms to those found in other beta-hemolytic streptococci. Both *erm* genes, encoding methylases which modify the binding site of these antibiotics in the ribosome and confer broad-spectrum resistance to macrolides, lincosamides and streptogramin B (MLS_B) antibiotics, and *mef* genes, encoding a pump which promotes the active efflux of the antibiotic, were detected in SDE (203, 206, 209, 210, 213). The prevailing mechanisms of resistance differed between studies (209, 210) and resistance rates to erythromycin in SDE have ranged from values below 10% (203, 209) up to 26% (141) in European countries and almost 30% in invasive isolates from the United States (17). Such variations are not unexpected, as it is known that the prevalence of the distinct macrolide resistance mechanisms can change both geographically and temporally within streptococcal species (212).

High resistance rates to tetracycline (30 to 40% and higher) were previously reported in GCGS isolates from several geographic locations (135, 141, 199), in line with the wide dissemination of tetracycline resistance determinants in many Gram positive bacteria (208). Tetracyclines inhibit protein synthesis by binding reversibly to the 30S ribosomal subunit and blocking the binding of the tRNA, and resistance to these agents arises upon acquisition of *tet* genes (208). At least three distinct genes encoding ribosomal protection proteins, *tet*(M), *tet*(O) and *tet*(S), and the *tet*(L) gene encoding an efflux pump were previously detected in SDE (141, 202, 207).

Aminoglycosides are another antimicrobial class interfering with protein synthesis by binding to the 30S ribosomal subunit (214). Although streptococci are intrinsically resistant to low concentrations of aminoglycosides (214), high-level resistance to these antimicrobial agents has been found to be confined to a few GCGS isolates (135). A

study by Galimand *et al.* (204), which characterized the genetic element responsible for high-level resistance to aminoglycosides in a group G human isolate, showed that it resulted from a recombination event between two widespread transposons found in staphylococci carrying the genetic determinants for three distinct aminoglycoside modifying enzymes.

Fluoroquinolones are not first-line options in the treatment of GCGS and other beta-hemolytic streptococci, but these agents retain excellent activity against these bacteria (199). Fluoroquinolones act by inhibiting the bacterial DNA gyrase and DNA topoisomerase IV (215). Alteration of the target enzymes is the more widespread mechanism of resistance in streptococci and confers high-level resistance to these antibiotics (215). Resistance rates not exceeding 1% have been reported in SDE isolates from the United States (17, 199) and Japan (122), but the potential emergence of resistance to fluoroquinolones in streptococci is a cause of concern. Non-susceptibility to this antimicrobial class has been reported to be increasing in GAS isolates from Belgium (216) and horizontal transfer events involving *S. dysgalactiae* have been proposed to contribute to the emergence of fluoroquinolone resistant-GAS (217).

4. TYPING METHODOLOGIES

The use of typing methods to determine the clonal relationships between distinct bacterial isolates enables the spread of defined clones and the population dynamics of bacteria to be studied. The information generated can therefore be used to derive hypothesis related to taxonomy, ecology and the study of pathogenesis (218). A practical example of these aspects is obtained by looking at how the distinct typing methodologies were used to characterize GCGS strains over the years. Both phenotypic and genotypic methods were essential to recognize the several GCGS populations which were ultimately allocated to distinct species.

4.1. Phenotypic methods

Phenotypic testing, *i.e.* to discriminate between isolates according to observable traits, has long been used to characterize GCGS. In addition to the hallmark of these methods, the Lancefield serologic technique (2), the epidemiologic characterization of GCGS strains involved biochemical testing (100), antimicrobial typing (197), serotyping (192), and also phage and bacteriocin typing (3). Biochemical testing, comprising both individual tests and commercially available panels (including enzymatic and sugar fermentation tests), was probably the most used (33).

In the particular case of SDE, the intraspecific characterization of human strains was also achieved by employing the serological technique known as T typing, which is widely used in GAS epidemiology (137). This method was applied to SDE in the studies conducted by Efstratiou *et al.* in the United Kingdom during the 1980's (110, 192, 219). T typing was used to follow healthcare-associated outbreaks by this bacterium and it was shown that more than 80% of the SDE isolates could be typed with this scheme, with minimal overlap with GAS T types (110). Although a few more examples exist in the GCGS literature reporting the use of this technique, its application did not stand to the present and is now in disuse.

Phenotypic typing methods are often insufficient for achieving the desirable degree of differentiation between individual isolates. Although a premise for these methods is that a phenotype results and accurately reflects a given genotype, this is not always the case, and genetic exchange may blur the evolutionary history suggested by a particular phenotype (218). These aspects, as well as reproducibility issues, make these methods

generally unsuitable for comprehensive studies of bacterial population structure and dynamics (218, 220).

The multilocus enzyme electrophoresis technique constitutes an exception within phenotypic methods, as it presents a performance similar to many genotypic methods. It measures the different electrophoretic mobility of a set of proteins to indirectly detect the allelic variation of the genes that encode those proteins (220), and it can be considered the precursor of the multilocus sequence typing (MLST) technique. Multilocus enzyme electrophoresis was used in the past as a reference method for defining the phylogenetic structure of clonal lineages in bacterial populations (218). In the specific case of GCGS, it was useful in characterizing the subspecies level divisions recognized nowadays for some of the members of the group (33, 83), including the current taxonomic status of SDE.

4.2. Genotypic methods

Genotyping allows the variation in the genomes of bacterial isolates to be assessed with respect to their composition, overall structure, or specific nucleotide sequences (218). Among the genotypic methods used over the years to type GCGS strains are included those based on the electrophoretic separation of DNA fragments, as is the case of the pulsed-field gel electrophoresis (PFGE) technique (46, 63), and those relying on the sequence analysis of one or more genes (97, 157). While the first are more adequate for short-term epidemiology, such as the follow-up of the increasing prevalence of a strain during an outbreak, some of the latter have the advantage of providing information on long-term evolutionary changes (220).

4.2.1. Pulsed-field gel electrophoresis (PFGE)

Although several methods of DNA profiling analysis were employed in the past to characterize GCGS, PFGE was the most widely used (46, 58, 63, 115, 124). This technique has a higher discriminatory power than many other methods (221), and it was shown to be more discriminative than other techniques for typing beta-hemolytic streptococci, including GCGS (222). Differentiation of human and animal strains of *S. dysgalactiae* is also accomplished by PFGE (46).

The PFGE technique involves the embedment of bacterial cells in agarose plugs, followed by cellular lysis *in situ*, therefore maintaining DNA integrity. The enzymatic

hydrolysis of the total DNA is then carried out with endonucleases that cleave infrequently in the genome. The agarose plugs containing the DNA are then inserted into the wells of an agarose gel. In contrast to conventional electrophoresis, higher voltages and a pattern of electric pulses in which the direction of current is switched are applied, allowing the separation of the large DNA fragments. The end result of this electrophoresis is the production of a profile pattern for each isolate. The PFGE profiles of the isolates are then compared with each other to evaluate their relatedness (221) (Fig. 1.4).

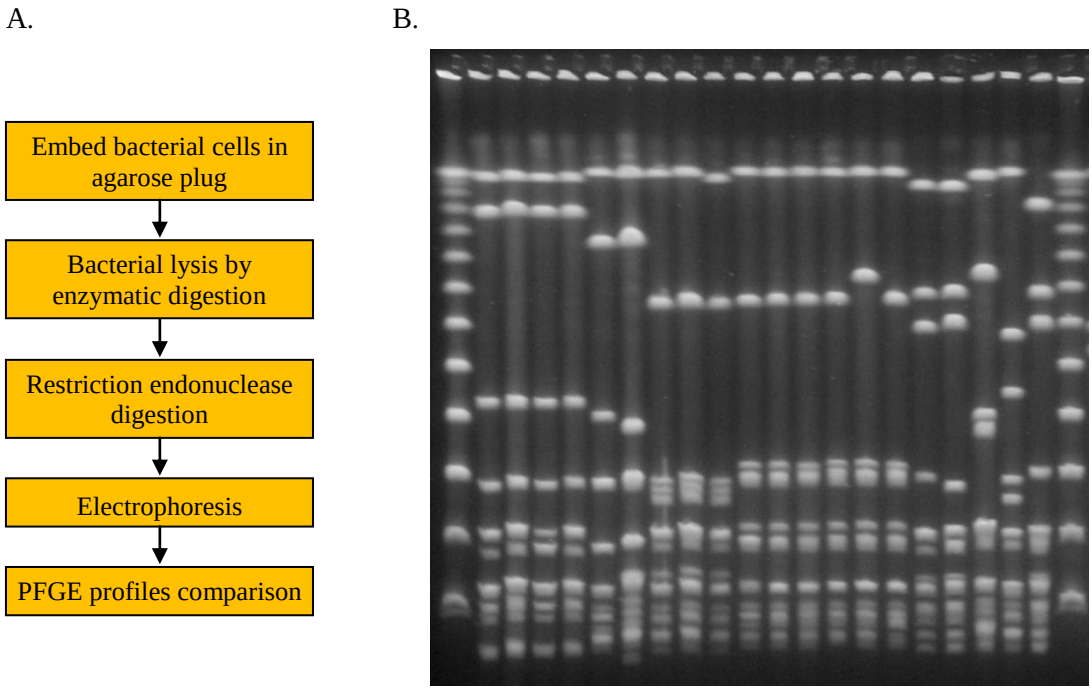

FIGURE 1.4. Illustration of the PFGE technique

(A) Schematic representation of the PFGE procedure. (B) PFGE profiles generated after SmaI digestion of total DNA from SDE isolates. The first and last lanes correspond to the lambda ladder PFGE marker (New England Biolabs, Beverly, MA).

The methods used to compare the PFGE profiles of each isolate have critical importance in the ability of the technique to ascertain relatedness. Guidelines for the visual interpretation of DNA patterns were proposed in 1995 by Tenover *et al.* (223). However, these criteria were proposed to be used with a small set of isolates, in the context of healthcare-associated outbreaks of limited duration (223). Visual comparison is not practical for studies comprising large populations of microorganisms collected over extended periods of time. In these cases, computer assisted analysis is useful.

Commercially available software can be used to construct a dendrogram, *i.e.*, a binary tree illustrating a cluster analysis performed on a number of isolates for a given typing data, based on pairwise comparisons among the isolates and depicting possible relationships between them (218).

The observable changes in PFGE patterns of related strains occur by random genetic events, including point mutations, insertions and deletions of DNA (223). PFGE analysis is not directed towards a specific region of the genome, but screens multiple *loci*, including uncharacterized regions that might accumulate increased variation in the bacterial population. Therefore, this technique allows detecting the microvariation that is generated in bacterial populations that circulate in a given geographic area and is especially useful in monitoring outbreaks.

4.2.2. *emm* typing

The *emm* typing technique was developed for the epidemiologic characterization of GAS isolates (158), to replace the successful use of the surface exposed M protein for serological typing of this species. A portion of the *emm* gene is sequenced, including the 5' region encoding the N-terminus of the mature protein, which determines the serological specificity. As observed for GAS, the *emm locus* is universally present in the SDE genome, as an *emm* gene can be amplified from virtually all SDE isolates analyzed (7, 17), and the *emm* alleles found in SDE also show polymorphisms at their 5' end (157). Thus, *emm* typing was successfully extended to the latter species. Distinct *emm* types are found in the two species (66, 135), although some common *emm* types have been occasionally detected (162).

The *emm* typing technique has become the most widely used typing scheme for the characterization of SDE isolates recovered in distinct parts of the world (7, 17, 106, 119, 140, 165). Characterization of SDE isolates recovered from both invasive and non-invasive infections showed the presence of multiple *emm* types (7, 106, 135), allowing SDE strains to be differentiated according to this marker. Over seventy distinct *emm* types are currently recognized for GCGS isolates, the vast majority corresponding to SDE *emm* types (http://www.cdc.gov/ncidod/biotech/strep/strepblast.htm). A few *emm* types have been reported from *S. canis*, but most of the isolates are PCR negative in the conventional *emm* typing scheme, limiting the usefulness of the technique to this bacterium. *emm* typing is not used to type other GCGS species, since no *emm* gene will

be amplified from their genome. In *S. equi* isolates, distinct M-like proteins were proposed as epidemiologic markers (224, 225).

The utilization of a typing method to compare groups of isolates presenting distinct virulence characteristics permits pathogenesis related markers to be identified (218). Given that the M protein is likely a major contributor to SDE virulence (as discussed above), the information obtained from *emm* typing may be clinically and epidemiologically relevant to ascertain differences in the virulence of SDE isolates, allowing particularly virulent strains or emerging clones to be detected. On the other hand, this technique might be inappropriate to infer evolutionary information. Since the *emm* gene encodes a surface exposed protein, this *locus* is likely evolving quickly due to the diversifying selection pressure that is exerted by the host immune response. Strains sharing a recent common ancestor might not be recognized as such, as the generation of variation may exceed the speed of spread (218). Moreover, results suggestive of recombination between SDE *emm* types have been detected (140), which may imply an additional loss of phylogenetic signal.

4.2.3. Multilocus sequence typing (MLST)

The MLST technique was developed in 1998 by Maiden *et al.* (226) and consists in sequencing internal fragments from a set of housekeeping genes, *i.e.* genes responsible for the maintenance of basic cellular functions. An allele number is arbitrarily assigned to the distinct nucleotide sequences obtained from each of the *loci* analyzed. By combining the allele numbers determined for a given isolate, an allelic profile is generated for each isolate which is named the MLST sequence type (ST) (Fig. 1.5).

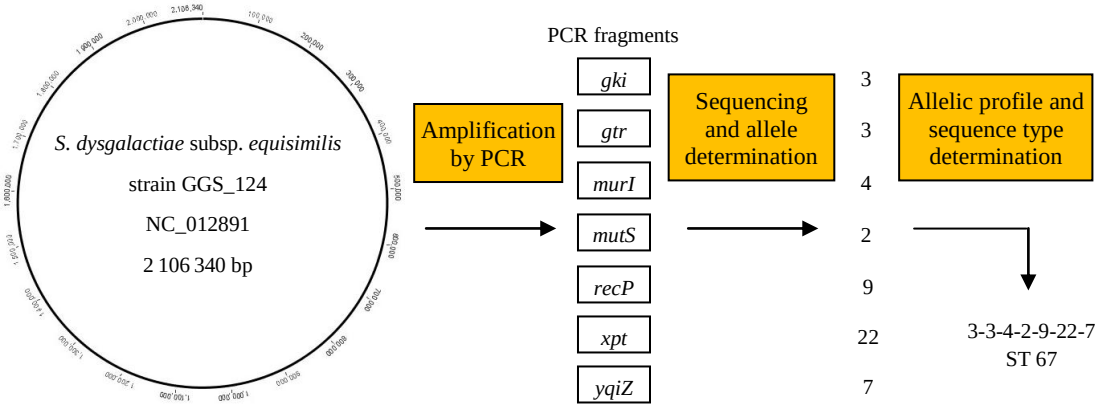

FIGURE 1.5. Illustration of the MLST technique

The allelic profile data is also used to establish the relationships between the distinct STs by grouping isolates that share the same alleles. By using clustering algorithms, such as eBURST (227) or goeBURST (228), isolates that share at least *n* -1 *loci* with each other are grouped together into clonal complexes (CCs). In the previous example, CCs are defined at the single-locus-variant (SLV) level, but analysis can also be conducted by considering more distantly related genotypes, such as double- and triple-locus-variants (DLVs and TLVs, respectively). The putative group founder in these CCs is identified as the ST which differs from the highest number of other genotypes in the CC at only one *locus*, *i.e.*, the ST with higher number of SLVs.

A MLST scheme was initially proposed for SDE in 2003, by adaption from the pre-existing scheme for GAS. However, this work could not be reproduced by other authors (44, 97), leading to the retraction of the original paper (229). Only in 2009 a novel MLST scheme for SDE was established (97). In the same work, this scheme was applied to the characterization of *S. canis*, although it was unsuccessful in typing all the isolates belonging to this species. A distinct MLST scheme is also available for the characterization of *S. equi* isolates (230).

One of the advantages of MLST over many other typing methods is the unambiguous and portable nature of the generated data, which allows it to be easily transferred and compared between laboratories (226). Another major feature of the method is that MLST analysis may be more relevant for the study of population genetics and dynamics than other methods used for epidemiologic purposes (220). By relying on housekeeping genes, MLST targets variation in bacterial genomes which is slowly accumulating in the population and that is expected to be selectively neutral (226). MLST sequence data can also be used to estimate recombination and mutation rates (231), making it possible to infer the importance of each if these forces in the emergence of clonal lineages, and to assess how commonly bacterial genomes undergo horizontal gene transfer events. Because of the aforementioned aspects, MLST has been extensively applied to many different bacterial species, providing vital information concerning their diversification and evolution.

4.2.4. Whole-genome sequencing

Whole-genome sequence methods offer the opportunity to assess genetic differences among members of a bacterial species or between unrelated bacteria with an increased

sensitivity in relation to the typing techniques described above. In particular, next generation sequencing technologies, which generate complete or nearly complete genome sequences of a great number of isolates with decreasing costs and turnaround times (232), allow the characterization of nucleotide changes accumulating in the entire genome which would be otherwise undetectable. Therefore, these technologies may have a strong impact if applied to epidemiological typing (such as surveillance and outbreak investigation) and routine genotypic identification and antimicrobial susceptibility testing (233), but has also the potential of considerably improving our knowledge on the virulence and evolution of bacterial pathogens.

Five SDE complete genome sequences are now available (January 2014), ranging from 2.08 to 2.18 Mb in size, while a single one has been released for *S. canis* (http://www.ncbi.nlm.nih.gov/genome). Among the five SDE genomes, including Lancefield group G (n = 3), C and A isolates, there is around a ninety percent overlap in sequence and most of the observable diversity within this set is due to the presence of prophages (234).

The first SDE whole-genome analysis, published for a group G isolate recovered from a patient with STSS (15), not only showed that it shared many virulence factor genes with GAS but also indicated the existence of related prophages in both genomes. In addition to the comparative analysis of gene content among these and other streptococci (172, 234), later studies have also evaluated the contribution of replacing and additive horizontal gene transfer events between the two species (235), the distribution of distinct types of mobile genetic elements (14, 15), and reconstructed the evolutionary history of specific *loci* (178). The evolutionary picture arising from the growing number of genomes being published for GCGS species reveals that strains belonging to the *S. dysgalactiae* (15, 172) and *S. canis* (14) *taxa* are more closely related to GAS (based on both genome wide and gene level comparisons, in nucleotide and amino acid sequences) and more distantly related to the members of the *S. equi taxon*.

5. REFERENCES

1. **Lancefield RC**. 1933. A serological differentiation of human and other groups of hemolytic streptococci. J. Exp. Med. **57**:571–595.

2. **Lancefield RC, Hare R**. 1935. The serological differentiation of pathogenic and non-pathogenic strains of hemolytic streptococci from parturient women. J. Exp. Med. **61**:335–349.

3. **Efstratiou A**. 1997. Pyogenic streptococci of Lancefield groups C and G as pathogens in man. Soc. Appl. Bacteriol. Symp. Ser. **26 Suppl**:72–79.

4. **Bradley SF, Gordon JJ, Baumgartner DD, Marasco WA, Kauffman CA**. 1991. Group C streptococcal bacteremia: analysis of 88 cases. Rev. Infect. Dis. **13**:270–280.

5. **Yuen KY, Seto WH, Choi CH, Ng W, Ho SW, Chau PY**. 1990. *Streptococcus zooepidemicus* (Lancefield group C) septicaemia in Hong Kong. J. Infect. **21**:241–250.

6. **Takahashi T, Ubukata K, Watanabe H**. 2011. Invasive infection caused by *Streptococcus dysgalactiae* subsp. *equisimilis*: characteristics of strains and clinical features. J. Infect. Chemother. **17**:1–10.

7. **Cohen-Poradosu R, Jaffe J, Lavi D, Grisariu-Greenzaid S, Nir-Paz R, Valinsky L, Dan-Goor M, Block C, Beall B, Moses AE**. 2004. Group G streptococcal bacteremia in Jerusalem. Emerging Infect. Dis. **10**:1455–1460.

8. **Gaunt N, Rogers K, Seal D, Denham M, Lewis J**. 1984. Necrotising fasciitis due to group C and G haemolytic streptococcus after chiropody. Lancet **1**:516.

9. **Wagner JG, Schlievert PM, Assimacopoulos AP, Stoehr JA, Carson PJ, Komadina K**. 1996. Acute group G streptococcal myositis associated with streptococcal toxic shock syndrome: case report and review. Clin. Infect. Dis. **23**:1159–1161.

10. **Wong SS, Lin YS, Mathew L, Rajagopal L, Sepkowitz D**. 2009. Increase in group G streptococcal infections in a community hospital, New York, USA. Emerging Infect. Dis. **15**:991–993.

11. **Sylvetsky N, Raveh D, Schlesinger Y, Rudensky B, Yinnon AM**. 2002. Bacteremia due to beta-hemolytic *Streptococcus* group G: increasing incidence and clinical characteristics of patients. Am. J. Med. **112**:622–626.

12. **Rantala S, Vuopio-Varkila J, Vuento R, Huhtala H, Syrjänen J**. 2009. Clinical presentations and epidemiology of β-haemolytic streptococcal bacteraemia: a population-based study. Clin. Microbiol. Infect. **15**:286–288.

13. **Barnham MRD, Weightman NC**. 2004. Changing incidence of detected streptococcal bacteraemia in North Yorkshire, England. Indian J. Med. Res. **119 Suppl**:160–163.

14. **Richards VP, Zadoks RN, Bitar PD, Lefébure T, Lang P, Werner B, Tikofsky L, Moroni P, Stanhope MJ**. 2012. Genome characterization and population genetic structure of the zoonotic pathogen, *Streptococcus canis*. BMC Microbiol. **12**:293.

15. **Shimomura Y, Okumura K, Murayama SY, Yagi J, Ubukata K, Kirikae T, Miyoshi-Akiyama T**. 2011. Complete genome sequencing and analysis of a Lancefield group G *Streptococcus dysgalactiae* subsp. *equisimilis* strain causing streptococcal toxic shock syndrome (STSS). BMC Genomics **12**:17.

16. **Beres SB, Sesso R, Pinto SWL, Hoe NP, Porcella SF, Deleo FR, Musser JM**. 2008. Genome sequence of a Lancefield group C *Streptococcus zooepidemicus* strain causing epidemic nephritis: new information about an old disease. PLoS ONE **3**:e3026.

17. **Broyles LN, Van Beneden C, Beall B, Facklam R, Shewmaker PL, Malpiedi P, Daily P, Reingold A, Farley MM**. 2009. Population-based study of invasive disease due to beta-hemolytic streptococci of groups other than A and B. Clin. Infect. Dis. **48**:706–712.

18. **Rantala S, Vuopio-Varkila J, Vuento R, Huhtala H, Syrjänen J**. 2009. Clinical presentations and epidemiology of beta-haemolytic streptococcal bacteraemia: a population-based study. Clin. Microbiol. Infect. **15**:286–288.

19. **Barnham M, Kerby J, Chandler RS, Millar MR**. 1989. Group C streptococci in human infection: a study of 308 isolates with clinical correlations. Epidemiol. Infect. **102**:379–390.

20. **Brandt CM, Spellerberg B**. 2009. Human infections due to *Streptococcus dysgalactiae* subspecies *equisimilis*. Clin. Infect. Dis. **49**:766–772.

21. **Whatmore AM, Engler KH, Gudmundsdottir G, Efstratiou A**. 2001. Identification of isolates of *Streptococcus canis* infecting humans. J. Clin. Microbiol. **39**:4196–4199.

22. **Ruoff KL, Whiley RA, Beighton D**. 2003. *Streptococcus*. p. 405-421. In P. R. Murray, E. J. Baron, J. H. Jorgensen, M. A. Pfaller and R. H. Yolken (ed.), Manual of Clinical Microbiology, 8th ed. American Society for Microbiology Press, Washington, D.C.

23. **Sherman JM**. 1937. The streptococci. Bacteriol. Rev. **1**:3–97.

24. **Humar D, Datta V, Bast DJ, Beall B, De Azavedo JCS, Nizet V**. 2002. Streptolysin S and necrotising infections produced by group G *Streptococcus*. Lancet **359**:124–129.

25. **Facklam R**. 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin. Microbiol. Rev. **15**:613–630.

26. **Lefébure T, Richards VP, Lang P, Pavinski-Bitar P, Stanhope MJ**. 2012. Gene repertoire evolution of *Streptococcus pyogenes* inferred from phylogenomic analysis with *Streptococcus canis* and *Streptococcus dysgalactiae*. PLoS ONE **7**:e37607.

27. **Vandamme P, Pot B, Falsen E, Kersters K, Devriese LA**. 1996. Taxonomic study of lancefield streptococcal groups C, G, and L (*Streptococcus dysgalactiae*) and proposal of *S. dysgalactiae* subsp. *equisimilis* subsp. nov. Int. J. Syst. Bacteriol. **46**:774–781.

28. **Woo PCY, Teng JLL, Lau SKP, Lum PNL, Leung K-W, Wong K-L, Li K-W, Lam K-C, Yuen K-Y**. 2003. Analysis of a viridans group strain reveals a case of bacteremia due to lancefield group G alpha-hemolytic *Streptococcus dysgalactiae* subsp *equisimilis* in a patient with pyomyositis and reactive arthritis. J. Clin. Microbiol. **41**:613–618.

29. **Dierksen KP, Tagg JR**. 2000. Haemolysin-deficient variants of *Streptococcus pyogenes* and *S. dysgalactiae* subsp. *equisimilis* may be overlooked as aetiological agents of pharyngitis. J. Med. Microbiol. **49**:811–816.

30. **Hitchcock CH**. 1924. Classification of the hemolytic streptococci by the precipitin reaction. J. Exp. Med. **40**:445–452.

31. **Krause RM**. 1963. Symposium on relationship of structure of microorganisms to their immunological properties. IV. Antigenic and biochemical composition of hemolytic streptococcal cell walls. Bacteriol. Rev. **27**:369–380.

32. **Bert F, Lambert-Zechovsky N**. 1997. Analysis of a case of recurrent bacteraemia due to group A *Streptococcus equisimilis* by pulsed-field gel electrophoresis. Infection **25**:250–251.

33. **Vieira VV, Teixeira LM, Zahner V, Momen H, Facklam RR, Steigerwalt AG, Brenner DJ, Castro AC**. 1998. Genetic relationships among the different phenotypes of *Streptococcus dysgalactiae* strains. Int. J. Syst. Bacteriol. **48**:1231–1243.

34. **Farrow JAE, Collins MD**. 1984. Taxonomic studies on streptococci of serological groups C, G and L and possibly related *taxa*. Syst. Appl. Microbiol. **5**:483–493.

35. **Devriese LA, Hommez J, Kilpper-Balz R, Schleifer K-H**. 1986. *Streptococcus canis* sp. nov.: A species of group G streptococci from animals. Int. J. Syst. Bacteriol. **36**:422–425.

36. **Galpérine T, Cazorla C, Blanchard E, Boineau F, Ragnaud J-M, Neau D**. 2007. *Streptococcus canis* infections in humans: retrospective study of 54 patients. J. Infect. **55**:23–26.

37. **Breiman RF, Silverblatt FJ**. 1986. Systemic *Streptococcus equi* infection in a horse handler – a case of human strangles. West. J. Med. **145**:385–386.

38. **Koh TH, Sng L-H, Yuen SM, Thomas CK, Tan PL, Tan SH, Wong NS**. 2009. Streptococcal cellulitis following preparation of fresh raw seafood. Zoonoses Public Health **56**:206–208.

39. **Marchandin H, Jumas-Bilak E, Boumzebra A, Vidal D, Jonquet O, Corne P**. 2007. Fatal *Streptococcus equi* subsp. *ruminatorum* infection in a man. Emerging Infect. Dis. **13**:1964–1966.

40. **Bert F, Lambert-Zechovsky N**. 1997. A case of bacteremia caused by *Streptococcus dysgalactiae*. Eur. J. Clin. Microbiol. Infect. Dis. **16**:324–326.

41. **Kawata K, Minakami T, Mori Y, Katsumi M, Kataoka Y, Ezawa A, Kikuchi N, Takahashi T**. 2003. rDNA sequence analyses of *Streptococcus dysgalactiae* subsp. *equisimilis* isolates from pigs. Int. J. Syst. Evol. Microbiol. **53**:1941–1946.

42. **Jensen A, Kilian M**. 2012. Delineation of *Streptococcus dysgalactiae*, its subspecies, and its clinical and phylogenetic relationship to *Streptococcus pyogenes*. J. Clin. Microbiol. **50**:113–126.

43. **Brandt CM, Haase G, Schnitzler N, Zbinden R, Lütticken R**. 1999. Characterization of blood culture isolates of *Streptococcus dysgalactiae* subsp. *equisimilis* possessing Lancefield's group A antigen. J. Clin. Microbiol. **37**:4194–4197.

44. **Tanaka D, Isobe J, Watahiki M, Nagai Y, Katsukawa C, Kawahara R, Endoh M, Okuno R, Kumagai N, Matsumoto M, Morikawa Y, Ikebe T, Watanabe H**. 2008. Genetic features of clinical isolates of *Streptococcus dysgalactiae* subsp. *equisimilis* possessing Lancefield's group A antigen. J. Clin. Microbiol. **46**:1526–1529.

45. **Erol E, Locke SJ, Donahoe JK, Mackin MA, Carter CN**. 2012. Beta-hemolytic *Streptococcus* spp. from horses: a retrospective study (2000-2010). J. Vet. Diagn. Invest. **24**:142–147.

46. **Bert F, Branger C, Poutrel B, Lambert-Zechovsky N**. 1997. Differentiation of human and animal strains of *Streptococcus dysgalactiae* by pulsed-field gel electrophoresis. FEMS Microbiol. Lett. **150**:107–112.

47. **Garvie EI, Farrow J a. E, Bramley AJ**. 1983. *Streptococcus dysgalactiae* (Diernhofer) nom. rev. Int. J. Syst. Bacteriol. **33**:404–405.

48. **Vela AI, Falsen E, Simarro I, Rollan E, Collins MD, Domínguez L, Fernandez-Garayzabal JF**. 2006. Neonatal mortality in puppies due to bacteremia by *Streptococcus dysgalactiae* subsp. *dysgalactiae*. J. Clin. Microbiol. **44**:666–668.

49. **Fernández-Aceñero MJ, Fernández-López P**. 2006. Cutaneous lesions associated with bacteremia by *Streptococcus dysgalactiae*. J. Am. Acad. Dermatol. **55 Suppl**:91–92.

50. **Park MJ, Eun I-S, Jung C-Y, Ko Y-C, Kim Y-J, Kim C-K, Kang E-J**. 2012. *Streptococcus dysgalactiae* subspecies *dysgalactiae* infection after total knee arthroplasty: a case report. Knee Surg. Relat. Res. **24**:120–123.

51. **Abdelsalam M, Chen S-C, Yoshida T**. 2010. Phenotypic and genetic characterizations of *Streptococcus dysgalactiae* strains isolated from fish collected in Japan and other Asian countries. FEMS Microbiol. Lett. **302**:32–38.

52. **Nomoto R, Munasinghe LI, Jin D-H, Shimahara Y, Yasuda H, Nakamura A, Misawa N, Itami T, Yoshida T**. 2004. Lancefield group C *Streptococcus dysgalactiae* infection responsible for fish mortalities in Japan. J. Fish. Dis. **27**:679–686.

53. **Nomoto R, Kagawa H, Yoshida T**. 2008. Partial sequencing of *sodA* gene and its application to identification of *Streptococcus dysgalactiae* subsp. *dysgalactiae* isolated from farmed fish. Lett. Appl. Microbiol. **46**:95–100.

54. **Netto LN, Leal CAG, Figueiredo HCP**. 2011. *Streptococcus dysgalactiae* as an agent of septicaemia in Nile tilapia, *Oreochromis niloticus* (L.). J. Fish Dis. **34**:251–254.

55. **Devriese LA, Cruz Colque JI, De Herdt P, Haesebrouck F**. 1992. Identification and composition of the tonsillar and anal enterococcal and streptococcal flora of dogs and cats. J. Appl. Bacteriol. **73**:421–425.

56. **Lyskova P, Vydrzalova M, Mazurova J**. 2007. Identification and antimicrobial susceptibility of bacteria and yeasts isolated from healthy dogs and dogs with otitis externa. J. Vet. Med. A Physiol. Pathol. Clin. Med. **54**:559–563.

57. **Murase T, Morita T, Sunagawa Y, Sawada M, Shimada A, Sato K, Hikasa Y**. 2003. Isolation of *Streptococcus canis* from a Japanese raccoon dog with fibrinous pleuropneumonia. Vet. Rec. **153**:471–472.

58. **DeWinter LM, Low DE, Prescott JF**. 1999. Virulence of *Streptococcus canis* from canine streptococcal toxic shock syndrome and necrotizing fasciitis. Vet. Microbiol. **70**:95–110.

59. **Pesavento PA, Bannasch MJ, Bachmann R, Byrne BA, Hurley KF**. 2007. Fatal *Streptococcus canis* infections in intensively housed shelter cats. Vet. Pathol. **44**:218–221.

60. **Lamm CG, Ferguson AC, Lehenbauer TW, Love BC**. 2010. Streptococcal infection in dogs: a retrospective study of 393 cases. Vet. Pathol. **47**:387–395.

61. **Chaffer M, Friedman S, Saran A, Younis A**. 2005. An outbreak of *Streptococcus canis* mastitis in a dairy herd in Israel. N. Z. Vet. J. **53**:261–264.

62. **Bert F, Lambert-Zechovsky N**. 1997. Septicemia caused by *Streptococcus canis* in a human. J. Clin. Microbiol. **35**:777–779.

63. **Takeda N, Kikuchi K, Asano R, Harada T, Totsuka K, Sumiyoshi T, Uchiyama T, Hosoda S**. 2001. Recurrent septicemia caused by *Streptococcus canis* after a dog bite. Scand. J. Infect. Dis. **33**:927–928.

64. **Lam MM, Clarridge JE, Young EJ, Mizuki S**. 2007. The other group G *Streptococcus*: increased detection of *Streptococcus canis* ulcer infections in dog owners. J. Clin. Microbiol. **45**:2327–2329.

65. **Woo PC, Fung AM, Lau SK, Wong SS, Yuen KY**. 2001. Group G beta-hemolytic streptococcal bacteremia characterized by 16S ribosomal RNA gene sequencing. J. Clin. Microbiol. **39**:3147–3155.

66. **Kittang BR, Skrede S, Langeland N, Haanshuus CG, Mylvaganam H**. 2011. *emm* gene diversity, superantigen gene profiles and presence of SlaA among clinical isolates of group A, C and G streptococci from western Norway. Eur. J. Clin. Microbiol. Infect. Dis. **30**:423–433.

67. **Siljander T, Karppelin M, Vähäkuopus S, Syrjänen J, Toropainen M, Kere J, Vuento R, Jussila T, Vuopio-Varkila J**. 2008. Acute bacterial, nonnecrotizing cellulitis in Finland: microbiological findings. Clin. Infect. Dis. **46**:855–861.

68. **Liao C-H, Liu L-C, Huang Y-T, Teng L-J, Hsueh P-R**. 2008. Bacteremia caused by group G streptococci, Taiwan. Emerging Infect. Dis. **14**:837–840.

69. **Jones FS**. 1919. The streptococci of equines. J. Exp. Med. **30**:159–178.

70. **Fernández E, Blume V, Garrido P, Collins MD, Mateos A, Domínguez L, Fernández-Garayzábal JF**. 2004. *Streptococcus equi* subsp. *ruminatorum* subsp. nov., isolated from mastitis in small ruminants. Int. J. Syst. Evol. Microbiol. **54**:2291–2296.

71. **Timoney JF**. 2004. The pathogenic equine streptococci. Vet. Res. **35**:397–409.

72. **Chalker VJ, Waller A, Webb K, Spearing E, Crosse P, Brownlie J, Erles K**. 2012. Genetic diversity of *Streptococcus equi* subsp. *zooepidemicus* and doxycycline resistance in kennelled dogs. J. Clin. Microbiol. **50**:2134–2136.

73. **Venn-Watson S, Daniels R, Smith C**. 2012. Thirty year retrospective evaluation of pneumonia in a bottlenose dolphin *Tursiops truncatus* population. Dis. Aquat. Org. **99**:237–242.

74. **Yamaguchi R, Nakamura S, Hori H, Kato Y, Une Y**. 2012. Purulent meningoventriculitis caused by *Streptococcus equi* subspecies *zooepidemicus* in a snow leopard (*Panthera uncia*). J. Comp. Pathol. **147**:397–400.

75. **Pesavento PA, Hurley KF, Bannasch MJ, Artiushin S, Timoney JF**. 2008. A clonal outbreak of acute fatal hemorrhagic pneumonia in intensively housed (shelter) dogs caused by *Streptococcus equi* subsp. *zooepidemicus*. Vet. Pathol. **45**:51–53.

76. **Lefort A, Lortholary O, Casassus P, Selton-Suty C, Guillevin L, Mainardi J-L, beta-Hemolytic Streptococci Infective Endocarditis Study Group**. 2002. Comparison between adult endocarditis due to beta-hemolytic streptococci (serogroups A, B, C, and G) and *Streptococcus milleri*: a multicenter study in France. Arch. Intern. Med. **162**:2450–2456.

77. **Eyre DW, Kenkre JS, Bowler ICJW, McBride SJ**. 2010. *Streptococcus equi* subspecies *zooepidemicus* meningitis – a case report and review of the literature. Eur. J. Clin. Microbiol. Infect. Dis. **29**:1459–1463.

78. **Hashikawa S, Iinuma Y, Furushita M, Ohkura T, Nada T, Torii K, Hasegawa T, Ohta M**. 2004. Characterization of group C and G streptococcal strains that cause streptococcal toxic shock syndrome. J. Clin. Microbiol. **42**:186–192.

79. **Korman TM, Boers A, Gooding TM, Curtis N, Visvanathan K**. 2004. Fatal case of toxic shock-like syndrome due to group C streptococcus associated with superantigen exotoxin. J. Clin. Microbiol. **42**:2866–2869.

80. **Duca E, Teodorovici G, Radu C, Vîţă A, Talaşman-Niculescu P, Bernescu E, Feldi C, Roşca V**. 1969. A new nephritogenic streptococcus. J. Hyg. **67**:691–698.

81. **Balter S, Benin A, Pinto SW, Teixeira LM, Alvim GG, Luna E, Jackson D, LaClaire L, Elliott J, Facklam R, Schuchat A**. 2000. Epidemic nephritis in Nova Serrana, Brazil. Lancet **355**:1776–1780.

82. **Waller AS, Paillot R, Timoney JF**. 2011. *Streptococcus equi*: a pathogen restricted to one host. J. Med. Microbiol. **60**:1231–1240.

83. **Jorm LR, Love DN, Bailey GD, McKay GM, Briscoe DA**. 1994. Genetic structure of populations of beta-haemolytic Lancefield group C streptococci from horses and their association with disease. Res. Vet. Sci. **57**:292–299.

84. **Goller KV, Fyumagwa RD, Nikolin V, East ML, Kilewo M, Speck S, Müller T, Matzke M, Wibbelt G**. 2010. Fatal canine distemper infection in a pack of African wild dogs in the Serengeti ecosystem, Tanzania. Vet. Microbiol. **146**:245–252.

85. **Speck S, Höner OP, Wachter B, Fickel J**. 2008. Characterization of *Streptococcus equi* subsp. *ruminatorum* isolated from spotted hyenas (*Crocuta crocuta*) and plains zebras (*Equus burchelli*), and identification of a M-like protein (SrM) encoding gene. Vet. Microbiol. **128**:148–159.

86. **Meyer A, Messer L, De Briel D, Moreau P**. 2011. Second reported case of human infection with *Streptococcus equi* subsp. *ruminatorum*. Joint Bone Spine **78**:303–305.

87. **Skaar I, Gaustad P, Tønjum T, Holm B, Stenwig H**. 1994. *Streptococcus phocae* sp. nov., a new species isolated from clinical specimens from seals. Int. J. Syst. Bacteriol. **44**:646–650.

88. **Vossen A, Abdulmawjood A, Lämmler C, Weiss R, Siebert U**. 2004. Identification and molecular characterization of beta-hemolytic streptococci isolated from harbor seals (*Phoca vitulina*) and grey seals (*Halichoerus grypus*) of the German North and Baltic Seas. J. Clin. Microbiol. **42**:469–473.

89. **Imai D, Jang S, Miller M, Conrad PA**. 2009. Characterization of beta-hemolytic streptococci isolated from southern sea otters (*Enhydra lutris nereis*) stranded along the California coast. Vet. Microbiol. **136**:378–381.

90. **Gibello A, Mata AI, Blanco MM, Casamayor A, Domínguez L, Fernández-Garayzabal JF**. 2005. First identification of *Streptococcus phocae* isolated from Atlantic salmon (*Salmo salar*). J. Clin. Microbiol. **43**:526–527.

91. **Romalde JL, Ravelo C, Valdés I, Magariños B, de la Fuente E, Martín CS, Avendaño-Herrera R, Toranzo AE**. 2008. *Streptococcus phocae*, an emerging pathogen for salmonid culture. Vet. Microbiol. **130**:198–207.

92. **Cherkaoui A, Emonet S, Fernandez J, Schorderet D, Schrenzel J**. 2011. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of beta-hemolytic streptococci. J. Clin. Microbiol. **49**:3004–3005.

93. **Anantha RV, Kasper KJ, Patterson KG, Zeppa JJ, Delport J, McCormick JK**. 2013. Fournier's gangrene of the penis caused by *Streptococcus dysgalactiae* subspecies *equisimilis*: case report and incidence study in a tertiary-care hospital. BMC Infect. Dis. **13**:381.

94. **Hassan AA, Khan IU, Abdulmawjood A, Lämmler C**. 2003. Development of PCR assays for detection of *Streptococcus canis*. FEMS Microbiol. Lett. **219**:209–214.

95. **McMillan DJ, Vu T, Bramhachari PV, Kaul SY, Bouvet A, Shaila MS, Karmarkar MG, Sriprakash KS**. 2010. Molecular markers for discriminating *Streptococcus pyogenes* and *S. dysgalactiae* subspecies equisimilis. Eur. J. Clin. Microbiol. Infect. Dis. **29**:585–589.

96. **Dawson ED, Taylor AW, Smagala JA, Rowlen KL**. 2009. Molecular detection of *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis*. Mol. Biotechnol. **42**:117–127.

97. **Ahmad Y, Gertz RE, Li Z, Sakota V, Broyles LN, Van Beneden C, Facklam R, Shewmaker PL, Reingold A, Farley MM, Beall BW**. 2009. Genetic relationships deduced from *emm* and multilocus sequence typing of invasive *Streptococcus dysgalactiae* subsp. *equisimilis* and *S. canis* recovered from isolates collected in the United States. J. Clin. Microbiol. **47**:2046–2054.

98. **Fox K, Turner J, Fox A**. 1993. Role of beta-hemolytic group C streptococci in pharyngitis: incidence and biochemical characteristics of *Streptococcus equisimilis* and *Streptococcus anginosus* in patients and healthy controls. J. Clin. Microbiol. **31**:804–807.

99. **Eriksson BK**. 1999. Anal colonization of group G beta-hemolytic streptococci in relapsing erysipelas of the lower extremity. Clin. Infect. Dis. **29**:1319–1320.

100. **Efstratiou A, Colman G, Hahn G, Timoney JF, Boeufgras JM, Monget D**. 1994. Biochemical differences among human and animal streptococci of Lancefield group C or group G. J. Med. Microbiol. **41**:145–148.

101. **Reid HF, Bassett DC, Poon-King T, Zabriskie JB, Read SE**. 1985. Group G streptococci in healthy school-children and in patients with glomerulonephritis in Trinidad. J. Hyg. **94**:61–68.

102. **Zaoutis T, Attia M, Gross R, Klein J**. 2004. The role of group C and group G streptococci in acute pharyngitis in children. Clin. Microbiol. Infect. **10**:37–40.

103. **Tartof SY, Farrimond F, de Matos JA, Reis JN, Ramos RTT, Andrade AN, dos Reis MG, Riley LW**. 2011. Inverse association between Lancefield group G *Streptococcus* colonization and sore throat in slum and nonslum settings in Brazil. J. Clin. Microbiol. **49**:409–412.

104. **McDonald M, Towers RJ, Andrews RM, Carapetis JR, Currie BJ**. 2007. Epidemiology of *Streptococcus dysgalactiae* subsp. *equisimilis* in tropical communities, Northern Australia. Emerging Infect. Dis. **13**:1694–1700.

105. **Haidan A, Talay SR, Rohde M, Sriprakash KS, Currie BJ, Chhatwal GS**. 2000. Pharyngeal carriage of group C and group G streptococci and acute rheumatic fever in an Aboriginal population. Lancet **356**:1167–1169.

106. **Bramhachari PV, Kaul SY, McMillan DJ, Shaila MS, Karmarkar MG, Sriprakash KS**. 2010. Disease burden due to *Streptococcus dysgalactiae* subsp. *equisimilis* (group G and C streptococcus) is higher than that due to *Streptococcus pyogenes* among Mumbai school children. J. Med. Microbiol. **59**:220–223.

107. **Semel JD, Goldin H**. 1996. Association of athlete's foot with cellulitis of the lower extremities: diagnostic value of bacterial cultures of ipsilateral interdigital space samples. Clin. Infect. Dis. **23**:1162–1164.

108. **Lewthwaite P, Parsons HK, Bates CJ, McKendrick MW, Dockrell DH**. 2002. Group G streptococcal bacteraemia: an opportunistic infection associated with immune senescence. Scand. J. Infect. Dis. **34**:83–87.

109. **Kristensen B, Schønheyder HC**. 1995. A 13-year survey of bacteraemia due to beta-haemolytic streptococci in a Danish county. J. Med. Microbiol. **43**:63–67.

110. **Efstratiou A**. 1989. Outbreaks of human infection caused by pyogenic streptococci of Lancefield groups C and G. J. Med. Microbiol. **29**:207–219.

111. **Goldman DA, Breton SJ**. 1978. Group C streptococcal surgical wound infections transmitted by an anorectal and nasal carrier. Pediatrics **61**:235–237.

112. **Schrieber L, Towers R, Muscatello G, Speare R**. 2013. Transmission of *Streptococcus dysgalactiae* subsp. *equisimilis* between child and dog in an aboriginal Australian community. Zoonoses Public Health. *In press*.

113. **Brouwer MC, Kasanmoentalib ES, Opstelten FWJ, van der Ende A, van de Beek D**. 2010. A horse bite to remember. Lancet **376**:1194.

114. **Aggarwal H, Smith AD, Kuo PC, Vikraman DS**. 2009. *Streptococcus equi* primary bacteremia in a hepatitis-C cirrhotic patient. Transplantation **87**:783.

115. **Abbott Y, Acke E, Khan S, Muldoon EG, Markey BK, Pinilla M, Leonard FC, Steward K, Waller A**. 2010. Zoonotic transmission of *Streptococcus equi* subsp. *zooepidemicus* from a dog to a handler. J. Med. Microbiol. **59**:120–123.

116. **Kuusi M, Lahti E, Virolainen A, Hatakka M, Vuento R, Rantala L, Vuopio-Varkila J, Seuna E, Karppelin M, Hakkinen M, Takkinen J, Gindonis V, Siponen K, Huotari K**. 2006. An outbreak of *Streptococcus equi* subspecies *zooepidemicus* associated with consumption of fresh goat cheese. BMC Infect. Dis. **6**:36.

117. **Francis AJ, Nimmo GR, Efstratiou A, Galanis V, Nuttall N**. 1993. Investigation of milk-borne *Streptococcus zooepidemicus* infection associated with glomerulonephritis in Australia. J. Infect. **27**:317–323.

118. **Kugu S, Sevim MS, Kaymak NZ, Erdogan G, Kandemir B, Dogan OK**. 2012. Exogenous group G *Streptococcus* endophthalmitis following intravitreal ranibizumab injection. Clin. Ophthalmol. **6**:1399–1402.

119. **Rantala S, Vahakuopus S, Vuopio-Varkila J, Vuento R, Syrjanen J**. 2010. *Streptococcus dysgalactiae* subsp. *equisimilis* bacteremia, Finland, 1995-2004. Emerging Infect. Dis. **16**:843–846.

120. **Carmeli Y, Ruoff KL**. 1995. Report of cases of and taxonomic considerations for large-colony-forming Lancefield group C streptococcal bacteremia. J. Clin. Microbiol. **33**:2114–2117.

121. **Ortel TL, Kallianos J, Gallis HA**. 1990. Group C streptococcal arthritis: case report and review. Rev. Infect. Dis. **12**:829–837.

122. **Takahashi T, Sunaoshi K, Sunakawa K, Fujishima S, Watanabe H, Ubukata K**. 2009. Clinical aspects of invasive infections with *Streptococcus dysgalactiae* ssp. *equisimilis* in Japan: differences with respect to *Streptococcus pyogenes* and *Streptococcus agalactiae* infections. Clin. Microbiol. Infect. **16**:1097-1103.

123. **Kittang BR, Langeland N, Skrede S, Mylvaganam H**. 2010. Two unusual cases of severe soft tissue infection caused by *Streptococcus dysgalactiae* subsp. *equisimilis*. J. Clin. Microbiol **48**:1484–1487.

124. **Ikebe T, Murayama S, Saitoh K, Yamai S, Suzuki R, Isobe J, Tanaka D, Katsukawa C, Tamaru A, Katayama A, Fujinaga Y, Hoashi K, Watanabe H, Working Group for Streptococci in Japan**. 2004. Surveillance of severe invasive group-G streptococcal infections and molecular typing of the isolates in Japan. Epidemiol. Infect. **132**:145–149.

125. **Ferrandière M, Cattier B, Dequin PF, Hazouard E, Legras A, Perrotin D**. 1998. Septicemia and meningitis due to *Streptococcus zooepidemicus*. Eur. J. Clin. Microbiol. Infect. Dis. **17**:290–291.

126. **Bordes-Benítez A, Sánchez-Oñoro M, Suárez-Bordón P, García-Rojas AJ, Saéz-Nieto JA, González-García A, Alamo-Antúnez I, Sánchez-Maroto A, Bolaños-Rivero M**. 2006. Outbreak of *Streptococcus equi* subsp. *zooepidemicus* infections on the island of Gran Canaria associated with the consumption of inadequately pasteurized cheese. Eur. J. Clin. Microbiol. Infect. Dis. **25**:242–246.

127. **Hugo-Persson M, Norlin K**. 1987. Erysipelas and group G streptococci. Infection **15**:184–187.

128. **Lewis RF, Balfour AE**. 1999. Group C streptococci isolated from throat swabs: a laboratory and clinical study. J. Clin. Pathol. **52**:264–266.

129. **Turner JC, Fox A, Fox K, Addy C, Garrison CZ, Herron B, Brunson C, Betcher G**. 1993. Role of group C beta-hemolytic streptococci in pharyngitis: epidemiologic study of clinical features associated with isolation of group C streptococci. J. Clin. Microbiol. **31**:808–811.

130. **Harrington AT, Clarridge JE 3rd**. 2013. Impact of identification of *Streptococcus dysgalactiae* subspecies *equisimilis* from throat cultures in an adult population. Diagn. Microbiol. Infect. Dis. **76**:20–23.

131. **Turner JC, Hayden FG, Lobo MC, Ramirez CE, Murren D**. 1997. Epidemiologic evidence for Lancefield group C beta-hemolytic streptococci as a cause of exudative pharyngitis in college students. J. Clin. Microbiol. **35**:1–4.

132. **Lindbaek M, Høiby EA, Lermark G, Steinsholt IM, Hjortdahl P**. 2005. Clinical symptoms and signs in sore throat patients with large colony variant beta-haemolytic streptococci groups C or G versus group A. Br. J. Gen. Pract. **55**:615–619.

133. **Benjamin JT, Perriello VA Jr**. 1976. Pharyngitis due to group C hemolytic streptococci in children. J. Pediatr. **89**:254–256.

134. **Nei T, Akutsu K, Shima A, Tsuboi I, Suzuki H, Yamamoto T, Tanaka K, Shinoyama A, Kojima Y, Washio Y, Okawa S, Sonobe K, Norose Y, Saito R**. 2012. A case of streptococcal toxic shock syndrome due to Group G streptococci identified as *Streptococcus dysgalactiae* subsp. *equisimilis*. J. Infect. Chemother. **18**:919-924.

135. **Lopardo HA, Vidal P, Sparo M, Jeric P, Centron D, Facklam RR, Paganini H, Pagniez NG, Lovgren M, Beall B**. 2005. Six-month multicenter study on invasive infections due to *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis* in Argentina. J. Clin. Microbiol. **43**:802–807.

136. **Ikebe T, Oguro Y, Ogata K, Katsukawa C, Isobe J, Shima T, Suzuki R, Ohya H, Tominaga K, Okuno R, Uchitani Y, Tada Y, Okabe N, Watanabe H**. 2010. Surveillance of severe invasive group G streptococcal infections in Japan during 2002-2008. Jpn. J. Infect. Dis. **63**:372–375.

137. **Cunningham MW**. 2000. Pathogenesis of group A streptococcal infections. Clin. Microbiol. Rev. **13**:470–511.

138. **McDonald MI, Towers RJ, Andrews RM, Benger N, Currie BJ, Carapetis JR**. 2006. Low rates of streptococcal pharyngitis and high rates of pyoderma in Australian aboriginal communities where acute rheumatic fever is hyperendemic. Clin. Infect. Dis. **43**:683–689.

139. **Reissmann S, Gillen CM, Fulde M, Bergmann R, Nerlich A, Rajkumari R, Brahmadathan KN, Chhatwal GS, Nitsche-Schmitz DP**. 2012. Region specific and worldwide distribution of collagen-binding M proteins with PARF motifs among human pathogenic streptococcal isolates. PLoS ONE **7**:e30122.

140. **Kittang BR, Skrede S, Langeland N, Haanshuus CG, Mylvaganam H**. 2011. *emm* gene diversity, superantigen gene profiles and presence of SlaA among clinical isolates of group A, C and G streptococci from western Norway. Eur. J. Clin. Microbiol. Infect. Dis. **30**:423–433.

141. **Loubinoux J, Plainvert C, Collobert G, Touak G, Bouvet A, Poyart C, CNR-Strep Network**. 2013. Adult invasive and non-invasive infections due to *Streptococcus dysgalactiae* subsp. *equisimilis* in France from 2006 to 2010. J. Clin. Microbiol. **51**:2724–2727.

142. **Reissmann S, Friedrichs C, Rajkumari R, Itzek A, Fulde M, Rodloff AC, Brahmadathan KN, Chhatwal GS, Nitsche-Schmitz DP**. 2010. Contribution of

Streptococcus anginosus to infections caused by groups C and G streptococci, southern India. Emerging Infect. Dis. **16**:656–663.

143. **Harris P, Siew D-A, Proud M, Buettner P, Norton R**. 2011. Bacteraemia caused by beta-haemolytic streptococci in North Queensland: changing trends over a 14-year period. Clin. Microbiol. Infect. **17**:1216–1222.

144. **Ramirez C, Turner J, Chance JF**. 2011. Non-group A streptococci as common isolates from throat culture among college students with pharyngitis. J. Am. Coll. Health **59**:237.

145. **Laupland KB, Ross T, Church DL, Gregson DB**. 2006. Population-based surveillance of invasive pyogenic streptococcal infection in a large Canadian region. Clin. Microbiol. Infect. **12**:224–230.

146. **Hindsholm M, Schønheyder HC**. 2002. Clinical presentation and outcome of bacteraemia caused by beta-haemolytic streptococci serogroup G. APMIS **110**:554–558.

147. **Myhre EB, Kronvall G**. 1977. Heterogeneity of nonimmune immunoglobulin Fc reactivity among gram-positive cocci: description of three major types of receptors for human immunoglobulin G. Infect. Immun. **17**:475–482.

148. **Myhre EB, Kuusela P**. 1983. Binding of human fibronectin to group A, C, and G streptococci. Infect. Immun. **40**:29–34.

149. **Fahnestock SR, Alexander P, Nagle J, Filpula D**. 1986. Gene for an immunoglobulin-binding protein from a group G streptococcus. J. Bacteriol. **167**:870–880.

150. **Collins CM, Kimura A, Bisno AL**. 1992. Group G streptococcal M protein exhibits structural features analogous to those of class I M protein of group A streptococci. Infect. Immun. **60**:3689–3696.

151. **Davies MR, McMillan DJ, Beiko RG, Barroso V, Geffers R, Sriprakash KS, Chhatwal GS**. 2007. Virulence profiling of *Streptococcus dysgalactiae* subspecies *equisimilis* isolated from infected humans reveals 2 distinct genetic lineages that do not segregate with their phenotypes or propensity to cause diseases. Clin. Infect. Dis. **44**:1442–1454.

152. **Vasi J, Frykberg L, Carlsson LE, Lindberg M, Guss B**. 2000. M-like proteins of *Streptococcus dysgalactiae*. Infect. Immun. **68**:294–302.

153. **Timoney JF, Artiushin SC, Boschwitz JS**. 1997. Comparison of the sequences and functions of *Streptococcus equi* M-like proteins SeM and SzPSe. Infect. Immun. **65**:3600–3605.

154. **Fischetti VA**. 1989. Streptococcal M protein: molecular design and biological behavior. Clin. Microbiol. Rev. **2**:285–314.

155. **Bisno AL, Craven DE, McCabe WR**. 1987. M proteins of group G streptococci isolated from bacteremic human infections. Infect. Immun. **55**:753–757.

156. **Jones KF, Fischetti VA**. 1987. Biological and immunochemical identity of M protein on group G streptococci with M protein on group A streptococci. Infect. Immun. **55**:502–506.

157. **Schnitzler N, Podbielski A, Baumgarten G, Mignon M, Kaufhold A**. 1995. M or M-like protein gene polymorphisms in human group G streptococci. J. Clin. Microbiol. **33**:356–363.

158. **Beall B, Facklam R, Thompson T**. 1996. Sequencing *emm*-specific PCR products for routine and accurate typing of group A streptococci. J. Clin. Microbiol. **34**:953–958.

159. **Steer AC, Magor G, Jenney AWJ, Kado J, Good MF, McMillan D, Batzloff M, Carapetis JR**. 2009. *emm* and C-repeat region molecular typing of beta-hemolytic streptococci in a tropical country: implications for vaccine development. J. Clin. Microbiol. **47**:2502–2509.

160. **Geyer A, Roth A, Vettermann S, Günther E, Groh A, Straube E, Schmidt K**. 1999. M protein of a *Streptococcus dysgalactiae* human wound isolate shows multiple binding to different plasma proteins and shares epitopes with keratin and human cartilage. FEMS Immunol. Med. Microbiol. **26**:11–24.

161. **Johansson HM, Mörgelin M, Frick I-M**. 2004. Protein FOG - a streptococcal inhibitor of neutrophil function. Microbiology **150**:4211–4221.

162. **Steer AC, Jenney AJW, Oppedisano F, Batzloff MR, Hartas J, Passmore J, Russell FM, Kado JHH, Carapetis JR**. 2008. High burden of invasive beta-haemolytic streptococcal infections in Fiji. Epidemiol. Infect. **136**:621–627.

163. **Bisno A, Brito M, Collins C**. 2003. Molecular basis of group A streptococcal virulence. Lancet Infect. Dis. **3**:191–200.

164. **Geyer A, Schmidt KH**. 2000. Genetic organisation of the M protein region in human isolates of group C and G streptococci: two types of multigene regulator-like (*mgrC*) regions. Mol. Gen. Genet. **262**:965–976.

165. **Tseng S-P, Lin Y-Y, Tsai J-C, Hsueh P-R, Chen H-J, Hung W-C, Teng L-J**. 2010. Distribution of *emm* types and genetic characterization of the *mgc* locus in group G *Streptococcus dysgalactiae* subsp. *equisimilis* from a hospital in northern Taiwan. J. Clin. Microbiol **48**:2975–2977.

166. **Egesten A, Frick I-M, Mörgelin M, Olin AI, Björck L**. 2011. Binding of albumin promotes bacterial survival at the epithelial surface. J. Biol. Chem. **286**:2469–2476.

167. **Linge HM, Sastalla I, Nitsche-Schmitz DP, Egesten A, Frick I-M**. 2007. Protein FOG is a moderate inducer of MIG/CXCL9, and group G streptococci are more tolerant than group A streptococci to this chemokine's antibacterial effect. Microbiology **153**:3800–3808.

168. **Nitsche DP, Johansson HM, Frick I-M, Mörgelin M**. 2006. Streptococcal protein FOG, a novel matrix adhesin interacting with collagen I in vivo. J. Biol. Chem. **281**:1670–1679.

169. **Nitsche-Schmitz DP, Johansson HM, Sastalla I, Reissmann S, Frick I-M, Chhatwal GS**. 2007. Group G streptococcal IgG binding molecules FOG and protein G have different impacts on opsonization by C1q. J. Biol. Chem. **282**:17530–17536.

170. **Hondorp ER, McIver KS**. 2007. The Mga virulence regulon: infection where the grass is greener. Mol. Microbiol. **66**:1056–1065.

171. **Panchaud A, Guy L, Collyn F, Haenni M, Nakata M, Podbielski A, Moreillon P, Roten C-AH**. 2009. M-protein and other intrinsic virulence factors of *Streptococcus pyogenes* are encoded on an ancient pathogenicity island. BMC Genomics **10**:198.

172. **Suzuki H, Lefébure T, Hubisz MJ, Pavinski Bitar P, Lang P, Siepel A, Stanhope MJ**. 2011. Comparative genomic analysis of the *Streptococcus dysgalactiae* species group: gene content, molecular adaptation, and promoter evolution. Genome Biol. Evol. **3**:168–185.

173. **Yang J, Liu Y, Xu J, Li B**. 2010. Characterization of a new protective antigen of *Streptococcus canis*. Vet. Res. Commun. **34**:413–421.

174. **Fulde M, Rohde M, Hitzmann A, Preissner KT, Nitsche‐Schmitz DP, Nerlich A, Chhatwal GS, Bergmann S**. 2011. SCM, a novel M-like protein from *Streptococcus canis*, binds (mini)-plasminogen with high affinity and facilitates bacterial transmigration. Biochem. J. **434**:523–535.

175. **Sachse S, Seidel P, Gerlach D, Günther E, Rödel J, Straube E, Schmidt K-H**. 2002. Superantigen-like gene(s) in human pathogenic *Streptococcus dysgalactiae*, subsp *equisimilis*: genomic localisation of the gene encoding streptococcal pyrogenic exotoxin G (*speG(dys)*). FEMS Immunol. Med. Microbiol. **34**:159–167.

176. **Korem M, Hidalgo-Grass C, Michael-Gayego A, Nir-Paz R, Salameh S, Moses AE**. 2013. Streptococcal pyrogenic exotoxin G gene in blood and pharyngeal isolates of *Streptococcus dysgalactiae* subspecies *equisimilis* has a limited role in pathogenesis. J. Microbiol. Immunol. Infect. *In press*.

177. **Brandt CM, Schweizer KG, Holland R, Lütticken R, Freyaldenhoven BS**. 2005. Lack of mitogenic activity of *speG*- and *speG(dys)*-positive *Streptococcus dysgalactiae* subspecies *equisimilis* isolates from patients with invasive infections. Int. J. Med. Microbiol. **295**:539–546.

178. **Okumura K, Shimomura Y, Yamagata Murayama S, Yagi J, Ubukata K, Kirikae T, Miyoshi-Akiyama T**. 2012. Evolutionary paths of streptococcal and staphylococcal superantigens. BMC Genomics **13**:404.

179. **Paillot R, Robinson C, Steward K, Wright N, Jourdan T, Butcher N, Heather Z, Waller AS**. 2010. Contribution of each of four superantigens to *Streptococcus equi*-induced mitogenicity, gamma interferon synthesis, and immunity. Infect. Immun. **78**:1728–1739.

180. **Zhao J, Hayashi T, Saarinen S, Papageorgiou AC, Kato H, Imanishi K, Kirikae T, Abe R, Uchiyama T, Miyoshi-Akiyama T**. 2007. Cloning, expression, and characterization of the superantigen streptococcal pyrogenic exotoxin G from *Streptococcus dysgalactiae*. Infect. Immun. **75**:1721–1729.

181. **Rato MG, Bexiga R, Nunes SF, Vilela CL, Santos-Sanches I**. 2010. Human group A streptococci virulence genes in bovine group C streptococci. Emerging Infect. Dis. **16**:116–119.

182. **Paillot R, Darby AC, Robinson C, Wright NL, Steward KF, Anderson E, Webb K, Holden MTG, Efstratiou A, Broughton K, Jolley KA, Priestnall SL, Marotti Campi MC, Hughes MA, Radford A, Erles K, Waller AS**. 2010. Identification of three novel superantigen-encoding genes in *Streptococcus equi* subsp. *zooepidemicus*, *szeF*, *szeN*, and *szeP*. Infect. Immun. **78**:4817–4827.

183. **Igwe EI, Shewmaker PL, Facklam RR, Farley MM, van Beneden C, Beall B**. 2003. Identification of superantigen genes *speM*, *ssa*, and *smeZ* in invasive strains of beta-

hemolytic group C and G streptococci recovered from humans. FEMS Microbiol. Lett. **229**:259–264.

184. **Ingrey KT, Ren J, Prescott JF**. 2003. A fluoroquinolone induces a novel mitogen-encoding bacteriophage in *Streptococcus canis*. Infect. Immun. **71**:3028–3033.

185. **Cleary PP, Peterson J, Chen C, Nelson C**. 1991. Virulent human strains of group G streptococci express a C5a peptidase enzyme similar to that produced by group A streptococci. Infect. Immun. **59**:2305–2310.

186. **Lindgren PE, Signäs C, Rantamäki L, Lindberg M**. 1994. A fibronectin-binding protein from *Streptococcus equisimilis*: characterization of the gene and identification of the binding domain. Vet. Microbiol. **41**:235–247.

187. **Gase K, Gase A, Schirmer H, Malke H**. 1996. Cloning, sequencing and functional overexpression of the *Streptococcus equisimilis* H46A *gapC* gene encoding a glyceraldehyde-3-phosphate dehydrogenase that also functions as a plasmin(ogen)-binding protein. Purification and biochemical characterization of the protein. Eur. J. Biochem. **239**:42–51.

188. **Malke H, Roe B, Ferretti JJ**. 1985. Nucleotide sequence of the streptokinase gene from *Streptococcus equisimilis* H46A. Gene **34**:357–362.

189. **Wolinowska R, Cegłowski P, Kok J, Venema G**. 1991. Isolation, sequence and expression in *Escherichia coli*, *Bacillus subtilis* and *Lactococcus lactis* of the DNase (streptodornase)-encoding gene from *Streptococcus equisimilis* H46A. Gene **106**:115–119.

190. **Steiner K, Malke H**. 2002. Dual control of streptokinase and streptolysin S production by the *covRS* and *fasCAX* two-component regulators in *Streptococcus dysgalactiae* subsp. *equisimilis*. Infect. Immun. **70**:3627–3636.

191. **Mora M, Bensi G, Capo S, Falugi F, Zingaretti C, Manetti AGO, Maggi T, Taddei AR, Grandi G, Telford JL**. 2005. Group A *Streptococcus* produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc. Natl. Acad. Sci. **102**:15641–15646.

192. **Efstratiou A**. 1983. The serotyping of hospital strains of streptococci belonging to Lancefield group C and group G. J. Hyg. **90**:71–80.

193. **Portnoy D, Prentis J, Richards GK**. 1981. Penicillin tolerance of human isolates of group C streptococci. Antimicrob. Agents Chemother. **20**:235–238.

194. **Pillai A, Thomas S, Williams C**. 2005. Clindamycin in the treatment of group G beta-haemolytic streptococcal infections. J. Infect. **51**:e207–211.

195. **CLSI**. Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Second Informational Supplement. CLSI document M100-S22. Wayne, PA: Clinical and Laboratory Standards Institute; 2012.

196. **Portnoy D, Wink I, Richards GK, Blanc MZ**. 1980. Bacterial endocarditis due to a penicillin-tolerant group C streptococcus. Can. Med. Assoc. J. **122**:69–70, 75.

197. **Zaoutis T, Schneider B, Steele Moore L, Klein JD**. 1999. Antibiotic susceptibilities of group C and group G streptococci isolated from patients with invasive infections:

evidence of vancomycin tolerance among group G serotypes. J. Clin. Microbiol. **37**:3380–3383.

198. **Portnoy D, Prentis J, Richards GK**. 1981. Penicillin tolerance of human isolates of group C streptococci. Antimicrob. Agents Chemother. **20**:235–238.

199. **Biedenbach DJ, Toleman MA, Walsh TR, Jones RN**. 2006. Characterization of fluoroquinolone-resistant beta-hemolytic *Streptococcus* spp. isolated in North America and Europe including the first report of fluoroquinolone-resistant *Streptococcus dysgalactiae* subspecies *equisimilis*: report from the SENTRY Antimicrobial Surveillance Program (1997-2004). Diagn. Microbiol. Infect. Dis. **55**:119–127.

200. **Safdar A, Rolston KVI**. 2006. Vancomycin tolerance, a potential mechanism for refractory gram-positive bacteremia observational study in patients with cancer. Cancer **106**:1815–1820.

201. **Savini V, Catavitello C, Talia M, Manna A, Pompetti F, Di Bonaventura G, Di Giuseppe N, Febbo F, Balbinot A, Di Zacomo S, Esattore F, D'Antonio D**. 2008. Beta-lactam failure in treatment of two group G *Streptococcus dysgalactiae* subsp. *equisimilis* pharyngitis patients. J. Clin. Microbiol. **46**:814–816.

202. **Jeric PE, Lopardo H, Vidal P, Arduino S, Fernandez A, Orman BE, Sordelli DO, Centrón D**. 2002. Multicenter study on spreading of the *tet*(M) gene in tetracycline-resistant *Streptococcus* group G and C isolates in Argentina. Antimicrob. Agents Chemother. **46**:239–241.

203. **Kataja J, Seppälä H, Skurnik M, Sarkkinen H, Huovinen P**. 1998. Different erythromycin resistance mechanisms in group C and group G streptococci. Antimicrob. Agents Chemother. **42**:1493–1494.

204. **Galimand M, Lambert T, Gerbaud G, Courvalin P**. 1999. High-level aminoglycoside resistance in the beta-hemolytic group G *Streptococcus* isolate BM2721. Antimicrob. Agents Chemother. **43**:3008–3010.

205. **Roberts MC**. 2008. Update on macrolide-lincosamide-streptogramin, ketolide, and oxazolidinone resistance genes. FEMS Microbiol. Lett. **282**:147–159.

206. **Palmieri C, Magi G, Creti R, Baldassarri L, Imperi M, Gherardi G, Facinelli B**. 2013. Interspecies mobilization of an *ermT*-carrying plasmid of *Streptococcus dysgalactiae* subsp. *equisimilis* by a coresident ICE of the ICESa2603 family. J. Antimicrob. Chemother. **68**:23–26.

207. **Liu L-C, Tsai J-C, Hsueh P-R, Tseng S-P, Hung W-C, Chen H-J, Teng L-J**. 2008. Identification of *tet*(S) gene area in tetracycline-resistant *Streptococcus dysgalactiae* subsp. *equisimilis* clinical isolates. J. Antimicrob. Chemother. **61**:453–455.

208. **Roberts MC**. 2005. Update on acquired tetracycline resistance genes. FEMS Microbiol. Lett. **245**:195–203.

209. **Amezaga MR, McKenzie H**. 2006. Molecular epidemiology of macrolide resistance in beta-haemolytic streptococci of Lancefield groups A, B, C and G and evidence for a new *mef* element in group G streptococci that carries allelic variants of *mef* and *msr*(D). J. Antimicrob. Chemother. **57**:443–449.

210. **Woo PCY, To APC, Tse H, Lau SKP, Yuen K-Y**. 2003. Clinical and molecular epidemiology of erythromycin-resistant beta-hemolytic lancefield group G streptococci causing bacteremia. J. Clin. Microbiol. **41**:5188–5191.

211. **Figueira-Coelho J, Ramirez M, Salgado MJ, Melo-Cristino J**. 2004. *Streptococcus agalactiae* in a large Portuguese teaching hospital: antimicrobial susceptibility, serotype distribution, and clonal analysis of macrolide-resistant isolates. Microb. Drug Resist. **10**:31–36.

212. **Silva-Costa C, Ramirez M, Melo-Cristino J**. 2005. Rapid inversion of the prevalences of macrolide resistance phenotypes paralleled by a diversification of T and emm types among *Streptococcus pyogenes* in Portugal. Antimicrob. Agents Chemother. **49**:2109–2111.

213. **Sunaoshi K, Murayama SY, Adachi K, Yagoshi M, Okuzumi K, Chiba N, Morozumi M, Ubukata K**. 2010. Molecular *emm* genotyping and antibiotic susceptibility of *Streptococcus dysgalactiae* subsp. *equisimilis* isolated from invasive and non-invasive infections. J. Med. Microbiol. **59**:82–88.

214. **Vakulenko SB, Mobashery S**. 2003. Versatility of aminoglycosides and prospects for their future. Clin. Microbiol. Rev. **16**:430–450.

215. **Hooper DC**. 2002. Fluoroquinolone resistance among Gram-positive cocci. Lancet Infect Dis **2**:530–538.

216. **Van Heirstraeten L, Leten G, Lammens C, Goossens H, Malhotra-Kumar S**. 2012. Increase in fluoroquinolone non-susceptibility among clinical *Streptococcus pyogenes* in Belgium during 2007-10. J. Antimicrob. Chemother. **67**:2602–2605.

217. **Pletz MWR, McGee L, Van Beneden CA, Petit S, Bardsley M, Barlow M, Klugman KP**. 2006. Fluoroquinolone resistance in invasive *Streptococcus pyogenes* isolates due to spontaneous mutation and horizontal gene transfer. Antimicrob. Agents Chemother. **50**:943–948.

218. **Van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S, Struelens M**. 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin. Microbiol. Infect. **13 Suppl 3**:1–46.

219. **Efstratiou A, Teare EL, McGhie D, Colman G**. 1989. The presence of M proteins in outbreak strains of *Streptococcus equisimilis* T-type 204. J. Infect. **19**:105–111.

220. **Feil EJ, Enright MC**. 2004. Analyses of clonality and the evolution of bacterial pathogens. Curr. Opin. Microbiol. **7**:308–313.

221. **Olive DM, Bean P**. 1999. Principles and applications of methods for DNA-based typing of microbial organisms. J. Clin. Microbiol. **37**:1661–1669.

222. **Bert F, Branger C, Lambert-Zechovsky N**. 1997. Pulsed-field gel electrophoresis is more discriminating than multilocus enzyme electrophoresis and random amplified polymorphic DNA analysis for typing pyogenic streptococci. Curr. Microbiol. **34**:226–229.

223. **Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, Swaminathan B**. 1995. Interpreting chromosomal DNA restriction patterns produced by

pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. **33**:2233–2239.

224. **Nicholson ML, Ferdinand L, Sampson JS, Benin A, Balter S, Pinto SW, Dowell SF, Facklam RR, Carlone GM, Beall B**. 2000. Analysis of immunoreactivity to a *Streptococcus equi* subsp. *zooepidemicus* M-like protein to confirm an outbreak of poststreptococcal glomerulonephritis, and sequences of M-like proteins from isolates obtained from different host species. J. Clin. Microbiol. **38**:4126–4130.

225. **Anzai T, Kuwamoto Y, Wada R, Sugita S, Kakuda T, Takai S, Higuchi T, Timoney JF**. 2005. Variation in the N-terminal region of an M-like protein of *Streptococcus equi* and evaluation of its potential as a tool in epidemiologic studies. Am. J. Vet. Res. **66**:2167–2171.

226. **Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG**. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. **95**:3140–3145.

227. **Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG**. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186**:1518–1530.

228. **Francisco AP, Bugalho M, Ramirez M, Carriço JA**. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics **10**:152.

229. 2009. Retraction. Directional gene movement from human-pathogenic to commensal-like streptococci. Infect. Immun. **77**:4688.

230. **Webb K, Jolley K, Mitchell Z, Robinson C, Newton J, Maiden M, Waller A**. 2008. Development of an unambiguous and discriminatory multilocus sequence typing scheme for the *Streptococcus zooepidemicus* group. Microbiology **154**:3016–3024.

231. **Feil EJ, Maiden MC, Achtman M, Spratt BG**. 1999. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. Mol. Biol. Evol. **16**:1496–1502.

232. **Dunne WM Jr, Westblade LF, Ford B**. 2012. Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. Eur. J. Clin. Microbiol. Infect. Dis. **31**:1719–1726.

233. **Köser CU, Ellington MJ, Cartwright EJP, Gillespie SH, Brown NM, Farrington M, Holden MTG, Dougan G, Bentley SD, Parkhill J, Peacock SJ**. 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog. **8**:e1002824.

234. **Watanabe S, Kirikae T, Miyoshi-Akiyama T**. 2013. Complete genome sequence of *Streptococcus dysgalactiae* subsp. *equisimilis* 167 carrying Lancefield group C antigen and comparative genomics of *S. dysgalactiae* subsp. *equisimilis* strains. Genome Biol. Evol. **5**:1644–1651.

235. **Choi SC, Rasmussen MD, Hubisz MJ, Gronau I, Stanhope MJ, Siepel A**. 2012. Replacing and additive horizontal gene transfer in *Streptococcus*. Mol. Biol. Evol. **29**:3309-3320.

AIMS OF THE THESIS

The increasing relevance of GCGS as agents of human infection contrasts with the limited knowledge on the population biology of these bacteria at the present. A rising number of reports are showing that these streptococci have genetic and clinical features analogous to some of the main pathogens of the genus, strengthening the view that GCGS may have a substantial pathogenic role in the human host. Even so, a limited number of studies have comprehensively focused in the epidemiology of human infections caused by GCGS, when compared to other species. In Portugal, as in many other geographic regions, there is no study assessing the clinical importance of this streptococcal group. Thus, the general purpose of this thesis was to obtain new insights into distinct aspects of the biology and epidemiology of GCGS infections in the human host, with particular emphasis on the characterization of the population clonal structure. Given that SDE was the most common GCGS species found causing human infections in Portugal, most of the work centered in this bacterium.

Particular emphasis was put on the following aspects:

- Evaluate the prevalence of the distinct GCGS species causing human infection in Portugal, allowing the comparison with other geographical areas and clarifying the significance of each species as human pathogens.
- Perform the epidemiologic characterization of the isolates using molecular typing techniques such as PFGE, *emm* typing and MLST, in order to determine the clonal diversity and structure of GCGS populations.
- Assess the congruence between the distinct typing techniques used, with the purpose of evaluating the ability of each of the methods for delineating GCGS lineages.
- Compare the clonal relationships between invasive (*i.e.* recovered from normally sterile sites) and non-invasive SDE isolates, in order to identify genetic lineages with enhanced virulence.
- Investigate the clonal relationship between SDE strains bearing Lancefield group C and G antigens.

- Develop a MLST protocol for SDE which allowed the clonal identity of PFGE and *emm* typing defined groups to be confirmed, as well as to produce a global view of the population structure of this pathogen.
- Determine the antimicrobial susceptibility of GCGS isolates to antimicrobial agents with potential use in the treatment of GCGS infections, and evaluate the implications of the detected resistance to the therapy of GCGS infections.
- Evaluate the molecular mechanisms of resistance to fluoroquinolones which could justify the high resistance rate to these agents identified in the current work for SDE isolates recovered in Portugal.
- Estimate the extension of genetic recombination between SDE and GAS in the genes defining fluoroquinolone resistance and the contribution of such events to the development of resistance.
- Extend the analysis of horizontal gene transfer events to the other *loci* analyzed in the current thesis, in order to evaluate how relevant such events could be in the evolution of GCGS.
- Develop a fully applicable MLST scheme for *S. canis*, allowing the clonal structure of the population of this species to be defined.
- Explore the zoonotic potential of distinct *S. canis* genetic lineages, by directly comparing isolates from different hosts.
- Establish MLST databases for both SDE and *S. canis*, providing new tools for the molecular typing of these microorganisms and improving our understanding of their global epidemiology.

CHAPTER 2

Clonal relationships between invasive and noninvasive Lancefield group C and G streptococci and *emm*-specific differences in invasiveness

The contents of this chapter were presented in:

Pinho MD, Melo-Cristino J, Ramirez M

2006. Journal of Clinical Microbiology **44**:841-846

ABSTRACT

Lancefield group G and group C streptococci (GGS and GCS, respectively) are pathogens responsible for a number of life-threatening infections. A collection of 116 recent (1998 to 2004) invasive (n = 28) and noninvasive (n = 88) GGS and GCS clinical isolates from Portugal were characterized. All isolates were identified as *Streptococcus dysgalactiae* subsp. *equisimilis* (SDE) and characterized by *emm* typing and DNA macrorestriction profiling using pulsed-field gel electrophoresis (PFGE). *emm* typing revealed the presence of 22 distinct types, including 3 novel types. PFGE identified 14 clones with more than two isolates, but over half of the isolates were concentrated in 3 large clones. Individual clones and *emm* types showed a low level of association, since the majority of the clones included more than one *emm* type and the same *emm* type was found among diverse genetic backgrounds. Two *emm* types, *stG2078* and *stG10*, were significantly more frequent among invasive isolates, and another two, *stG6792* and *stG166b*, were present only in noninvasive isolates, suggesting a correlation between *emm* type and invasive disease potential.

INTRODUCTION

Beta-hemolytic, large-colony-forming (diameter, >0.5 mm) Lancefield group G and group C streptococci (GGS and GCS, respectively) are increasingly recognized as important human pathogens (6, 16, 23), but the study of these infections has been hampered by the heterogeneity of the species presenting Lancefield group C and G polysaccharides. The taxonomic classification of both GGS and GCS is in a state of flux, but at least three species have been found in human infections: *Streptococcus dysgalactiae* subsp. *equisimilis* (SDE), *Streptococcus equi* subsp. *zooepidemicus*, and *Streptococcus canis* (12). Characterization of the isolates recovered from human infections to the species level is uncommon, and most of the available data consider GGS and GCS independently, with more information available on GGS bacteremia (6, 27) and much less for GCS (5). Invasive GGS infections were mostly reported to be caused by SDE (6, 27). This species may also present group C polysaccharide, and such isolates have been recovered from human infections (26). However, few studies present species identification of GCS, and a study from Hong Kong reported that most group C beta-hemolytic streptococcal bacteremias are caused by *S. equi* subsp. *zooepidemicus* (28), while the contributions of the different species to GCS infections in other geographic locations remain mostly unknown.

Although both GGS and GCS are part of the normal human microbiota (9), they have also been identified as causative agents of infections of the respiratory tract (29), infections of the skin and soft tissue, and life threatening infections such as endocarditis, bacteremia, and meningitis, frequently with a poor prognosis (3, 9). More recently, an increasing number of reports described the association with streptococcal syndromes typically caused by *Streptococcus pyogenes* (Lancefield group A streptococcus [GAS]), such as streptococcal toxic shock syndrome (STSS) (16, 19) and acute rheumatic fever (15). Despite differences in the disease burden attributable to each of these organisms, GGS, GCS, and GAS are closely related genetically and share several virulence factors (9, 12). Proteins similar to the M protein (encoded by the *emm* gene), a key virulence factor in GAS, have long been identified in GGS and GCS from human infections, and the *emm* genes of both species share structural features such as substantial polymorphisms at the 5′ end (21). The M protein contributes considerably to the invasive capacity of GAS by mediating the antiphagocytic, adherence, and internalization processes (7). For GAS, different *emm* genes have been suggested as

genotypic markers for tissue site preference (20), and a recent study reevaluated an older observation that isolates expressing certain M proteins were more frequently associated with invasive disease (11). It is therefore conceivable that the M protein also plays an important part in GGS and GCS pathogenesis, although its role remains largely unexplored and most of the work has concentrated on the presence of superantigens in these bacteria (16, 19).

The M protein has also been used to differentiate strains, and the sequence-based typing scheme developed for GAS, relying on the hypervariable region of the *emm* gene, was successfully applied to GGS and GCS (http://www.cdc.gov/ncidod/biotech/strep/strepindex.htm). Other molecular typing techniques, such as macrorestriction profiling using pulsed-field gel electrophoresis (PFGE), have also been used successfully with pyogenic streptococci, including GGS and GCS (2).

The higher rate of isolation of these organisms in recent studies may be due to improved detection methods, an increase in virulence, or an expanding population of compromised hosts. In this study we have used *emm* typing and PFGE to characterize a collection of invasive (i.e., recovered from normally sterile sites) and noninvasive GGS and GCS isolated in Portugal during the years 1998 to 2004. The molecular characterization of these isolates allowed us to test whether PFGE-defined clones associate specifically with certain *emm* types and if there was a higher invasive disease potential of particular *emm* types and PFGE-defined clones.

MATERIALS AND METHODS

Bacterial isolates

A total of 116 beta-hemolytic, large-colony-forming (diameter, >0.5 mm) GGS and GCS isolates from clinical infections were collected from June 1998 to December 2004. The distribution of isolates over the study period was as follows: 5 in 1998, 14 in 1999, 12 in 2000, 14 in 2001, 14 in 2002, 35 in 2003, and 22 in 2004. The Lancefield groups and sources of the isolates are given in Table 2.1. The strains were recovered in nine laboratories, located in Portugal, that were asked to submit all nonduplicate GGS and GCS isolates associated with human infections.

TABLE 2.1. GGS and GCS isolates recovered from invasive and noninvasive infections during 1998 to 2004 in Portugal

Source	No. of isolates in Lancefield group:		Total no. of isolates
	C	G	
Blood	6	16	22
Other normally sterile fluids[a]	1	5	6
Pharyngeal exudate	8	18	26
Sputum	2	9	11
Skin and soft tissue	9	39	48
Vaginal exudate	0	2	2
Urine	0	1	1
Total	26	90	116

[a] Articular fluid (n = 2), peritoneal fluid (n = 2), or pleural fluid (n = 2).

Species identification

Strains were identified to the genus level by the submitting laboratory. Upon receipt, the Lancefield group was confirmed by a commercial latex agglutination technique, the Slidex Strepto-Kit (bioMérieux, Marcy-l'Etoile, France). All further studies were carried out at the Laboratory of Microbiology of the Lisbon School of Medicine. Beta-hemolysis and colony size were confirmed in tryptic soy agar (Oxoid, Hampshire, England) supplemented with 5% (vol/vol) defibrinated sheep blood, after overnight incubation at 37°C. Further identification to the species level was done using the API 20 Strep system (bioMérieux, Marcy-l'Etoile, France). Additionally, all the invasive isolates were subjected to 16S rRNA sequencing. Two generic primers for gram-positive bacteria (24) were used to amplify and sequence a 1,419-bp fragment of the

16S rRNA coding region. Template DNA for the PCR was prepared according to the method recommended for *emm* typing (http://www.cdc.gov/ncidod/biotech/strep/protocols.htm). One microliter of DNA template lysate was added to the PCR mixture containing 1× PCR buffer (100 mM Tris-HCl, 500 mM KCl, pH 9, 0.1% gelatin, 1% [vol/vol] Triton X-100), 200 μM deoxynucleoside triphosphates (MBI Fermentas, Vilnius, Lithuania), 0.5 μM primers, 1.5 mM $MgCl_2$, and 1 U of SUPER *Tth* DNA polymerase (HT Biotech, Cambridge, United Kingdom) in a final volume of 50 μl. PCR conditions for amplification were as follows: 35 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 2 min, and a final extension at 72°C for 10 min in a Biometra T gradient thermocycler (Goettingen, Germany). Amplification products were purified using a High Pure PCR product purification kit (Roche, Mannheim, Germany) and sequenced with the primers used for amplification. The sequences were compared to those deposited in the Ribosomal Database Project II (http://rdp.cme.msu.edu/) and GenBank.

***emm* typing**

Isolates were *emm* typed by following the protocols and recommendations for GAS of the Centers for Disease Control and Prevention (CDC) (http://www.cdc.gov/ncidod/biotech/strep/protocols.htm). Briefly, *emm* genes from GGS and GCS were amplified with primer 1 and primer 2 and sequenced with emmseq2. The first 240 bases of sequence were used to query the CDC streptococcal *emm* sequence database (http://www.cdc.gov/ncidod/biotech/strep/strepindex.htm). When a type assignment was not obtained, the full sequence was used to query GenBank. An *emm* type was defined as ≥95% identical over the first 160 bases of sequence obtained with primer emmseq2 (1). Sequences that did not show an identity to previously described sequences sufficient for *emm* type assignment were submitted to Bernard Beall at the CDC and were assigned an *emm* type designation.

PFGE macrorestriction profiling

Agarose plugs of total DNA were prepared according to methods described elsewhere (13), with some modifications. Briefly, cells were embedded in agarose plugs and lysed with a solution containing 5 U/ml mutanolysin, 1 mg/ml lysozyme, and 30 μg/ml RNase A (Sigma-Aldrich, Steinheim, Germany) for 5 h at 37°C and with 0.5

mg/ml proteinase K (Roche, Mannheim, Germany) for 17 h at 50°C. After digestion with SmaI (Fermentas, Vilnius, Lithuania), the fragments were resolved by PFGE as described previously (13).

PFGE patterns were compared by using Bionumerics software (Applied Maths, Sint-Martens-Latem, Belgium) to create dendrograms by the unweighted-pair group method with arithmetic means (UPGMA). The Dice similarity coefficient was used with optimization and position tolerance settings of 1.0 and 1.5, respectively. Clones were defined as groups of isolates ($n \geq 2$) presenting profiles ≥80% related on the dendrogram, as previously described for *Streptococcus pneumoniae* (22).

Estimation of the invasiveness of *emm* types and PFGE-defined clones

In order to compare the probability of invasive disease due to individual *emm* types and clones, an empirical odds ratio (OR) and 95% confidence intervals (CI) (8) were calculated by reference to all other *emm* types and clones as previously described (4). The OR was calculated as $(ad)/(bc)$, where a is the number of invasive A *emm* types or clones, b is the number of noninvasive A *emm* types or clones, c is the number of non-A *emm* types or clones, and d is the number of noninvasive non-A *emm* types or clones. It follows from the formula presented that it is not possible to calculate an OR when no isolates of a given *emm* type or clone were recovered from invasive infections.

The choice of using all other *emm* types and clones to measure the reference OR was supported by prior studies (reference 4 and references therein) that also provide a discussion of the strong points of this method. The ORs calculated in this way will be comparable between studies, although the set of all other strains that is used to measure the reference OR will be different. In contrast, the alternative hypotheses, such as choosing an arbitrary *emm* type or clone or using the pooled minor *emm* types or clones as a reference, suffer from more-severe pitfalls. The choice of a particular *emm* type or clone will cause the ORs to change significantly depending on which *emm* type or clone is chosen and will prevent comparison of the ORs between studies when that particular *emm* type or clone is not found. The pooling of minor types suffers from similar limitations, and both alternatives reduce statistical efficiency relative to that obtained by using all other *emm* types or clones, justifying the use of all other *emm* types and clones to measure the reference OR. An OR of 1 indicates that the *emm* type or clone was

equally likely to be invasive or noninvasive, whereas an OR of >1 or <1 indicated an increased or reduced probability to cause invasive disease, respectively.

RESULTS

Isolate identification

All 116 isolates were identified with the API 20 Strep system alone or together with 16S rRNA gene sequencing and were shown to be SDE independently of their Lancefield group polysaccharide. All but five strains could be identified to the species level by using the API 20 Strep system with a confidence in the identification above 95%, according to the manufacturer's literature. Each biochemical profile obtained with a confidence in the identification below 99% had a matching 16S rRNA gene sequence available, confirming the species identification. Only one isolate had an inconclusive identification with the API 20 Strep system (a group G blood isolate that fermented mannitol and sorbitol); however, the sequence of the 16S rRNA gene clearly identified it as SDE. Overall, 16 distinct biochemical profiles were obtained. Almost half of the strains (44%) fermented lactose, and 41% hydrolyzed esculin. Other tests for which results differed occasionally included pyrrolidone arylamidase and beta-galactosidase production.

The sequence of the 16S rRNA gene was obtained for all 28 invasive isolates; they differed at only 1 base in a 1,419-bp fragment of the gene and were all 99.9% identical to SDE ATCC 35666 (GenBank accession number AB096755).

emm type distribution and clonal relationships among isolates

emm genes were successfully amplified and sequenced for all the 116 SDE isolates analyzed in our study. We found 22 distinct *emm* types (Table 2.2), including 3 with novel *emm* sequences: *stGLP1*, *stGLP2*, and *stGLP3*. Overall, the eight most prevalent *emm* types accounted for approximately 75% of the isolates (Table 2.2). Most *emm* types were found exclusively among either group G or group C isolates. Exceptions occurred with *stC839* (nine group C isolates and one group G isolate) and *stG485* (six group C and three group G isolates).

Analysis of the PFGE patterns generated after digestion with SmaI revealed 14 clones differing greatly in size (Fig. 2.1). Over half of the isolates were distributed in three large clones, D_{33}, J_{19}, and F_{16} (the subscript number in each clone designation is the number of isolates) (Table 2.2). Clone D_{33} stands out, because it comprised almost one-third of the isolates analyzed. During the study period there were no clear trends

regarding the prevalences of the various *emm* types, and isolates representing the three major clones were recovered during all the years of the study and in most laboratories.

TABLE 2.2. *emm* types and distribution of invasive isolates among PFGE clones

***emm* type**[a]	**No. of isolates in group:**		**No. of invasive/noninvasive isolates in PFGE clone**[b]**:**								**OR (95% CI)**[d]
	C	**G**	**D_{33}**	**J_{19}**	**F_{16}**	**B_7**	**G_7**	**I_6**	**Other**[c]	**Total**	
stG10	0	16	7/9							7/9	2.9 (1.1-7.6)
stG6792	0	14			0/9				0/5	0/14	NA
stG6	0	12	1/0	0/5				0/2	1/3	2/10	NS
stG480	0	10	0/3			0/3		0/1	1/2	1/9	NS
stG2078	0	10	3/0	2/2				2/1		7/3	9.4 (2.2-41.2)
stC839	9	1	0/1		0/2		1/6			1/9	NS
stG485	6	3		2/4		1/0			1/1	4/5	NS
stG166b	0	8	0/6						0/2	0/8	NA
stG62647	4	0		1/2					0/1	1/3	NS
stG643	0	4		1/0		0/1			1/1	2/2	NS
stC36	3	0			2/0				0/1	2/1	NS
stC6979	3	0							0/3	0/3	NA
stC74a	0	2				0/2				0/2	NA
stG245	0	2	0/1						1/0	1/1	NS
stG652	0	2							0/2	0/2	NA
Other	1	6	0/2		0/3				0/2	0/7	NA
Total	26	90	11/22	6/13	2/14	1/6	1/6	2/4	5/23	28/88	NA

[a] *emm* types are given in decreasing order of frequency. Other *emm* types with a single representative isolate included *emm57*, *stG4974*, *stG5420*, *stCK401*, *stGLP1*, *stGLP2*, and *stGLP3*.

[b] PFGE clones were designated arbitrarily by capital letters and a subscript number indicating the number of isolates included in the clone.

[c] PFGE clones with <5 isolates. *stG6792* isolates were distributed among four distinct clones, *stG480* and *stC6979* isolates among three, *stG6*, *stG485*, *stG166b*, and *stG652* isolates among two, and *stG643* isolates formed a single clone. Five noninvasive isolates (three group G and two group C isolates) had unique PFGE patterns (i.e., shared less than 80% similarity with any other isolate in the collection).

[d] An OR of >1 indicates increased invasive disease potential. NS, not significant; NA, not applicable.

Half of the clones included both group G and C isolates; however, the largest clone (D_{33}) included no group C isolates, and more than half of GCS (*n* = 14) were concentrated in two clones – G_7 and J_{19} – of which G_7 included only GCS isolates. The majority of the clones included more than one *emm* type (Table 2.2). In D_{33} nine distinct *emm* types were found; however, *stG10* (*n* = 16) and *stG166b* (*n* = 6) together accounted for two-thirds of the isolates. Only clone G_7, already noted for including only

GCS isolates, was associated with a single *emm* type (*stC839*). Within clones, subgroups of the same *emm* type could be defined; still, there was no similarity cutoff that allowed the separation of all isolates presenting distinct *emm* types.

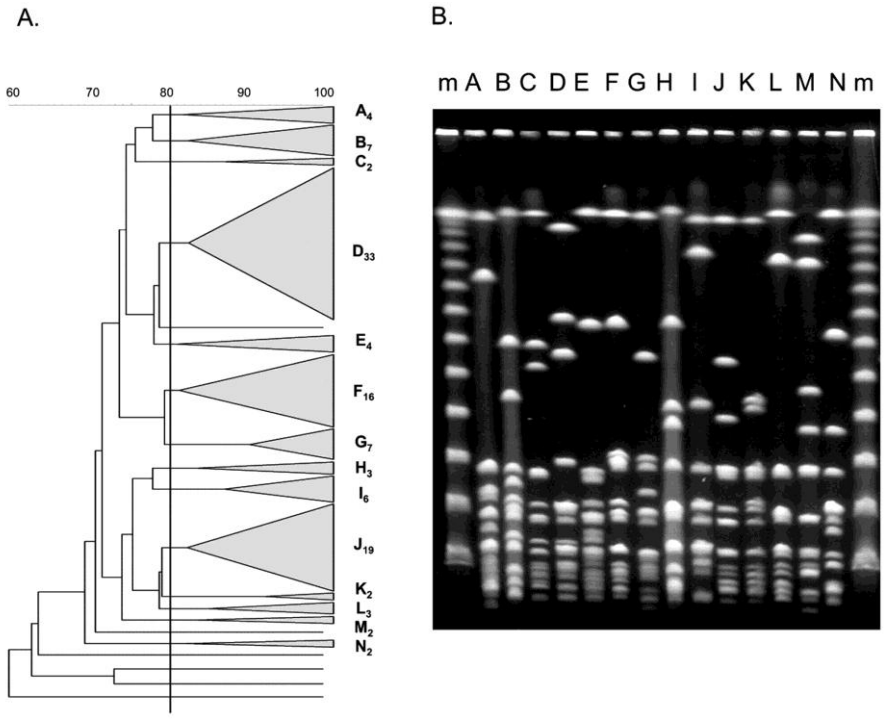

FIGURE 2.1. PFGE SmaI macrorestriction profile analysis of GGS and GCS isolates from invasive and noninvasive infections.

(A) Dendrogram showing UPGMA cluster analysis of the PFGE profiles of the 116 isolates studied. Dice coefficients (percentages) are indicated in the scale above the dendrogram. Each clone (defined as a group of ≥2 isolates with a Dice coefficient of ≥80%) is represented by a triangle proportional to the number of isolates included in the clone. Clones are designated by capital letters and a subscript number indicating the number of isolates included in the clone. (B) PFGE profiles of representatives of each clone. Capital letters above the lanes correspond to clone designations; m, lambda ladder PFGE marker (New England Biolabs, Beverly, MA).

With the exception of *stG10,* all the *emm* types had representatives in more than one clone, although the majority of the isolates of each *emm* type were represented in a single clone. In contrast, *stG480* and *stG2078* were highly dispersed among distinct genetic backgrounds.

Invasive isolates

None of the clones identified was significantly associated with invasive infection. However, it is noteworthy that a significant number of invasive isolates (*stG10* [n = 6] and *stG2078* [n = 3]) were tightly clustered within clone D_{33} at a similarity of 97.5% in the dendrogram.

ORs calculated to evaluate the invasiveness of the various *emm* types are presented in Table 2.2. The *emm* types *stG10* and *stG2078* were significantly more associated with invasive disease. While all the *stG10* isolates were present in a single clone, D_{33}, *stG2078* isolates were found to correspond to diverse genetic backgrounds (represented in three different clones). The *emm* types *stG6792* (*n*= 14) and *stG166b* (*n* = 8), each distributed in two clones, were found only among noninvasive isolates.

DISCUSSION

This study aimed to characterize human infections caused by GGS and GCS. Identification to the species level revealed that all isolates were SDE independently of their Lancefield polysaccharide group. This finding is not unique, since this species has been shown to express both C and G polysaccharides (26) and a recent study of GGS and GCS pharyngitis in the United States identified SDE as the sole species (29), but it was in sharp contrast with the situation in Hong Kong. In this region, SDE was the only species associated with GGS (27) but *S. equi* subsp. *zooepidemicus* was the only species identified in GCS bacteremia (28). In the latter study, the same strains were found in pigs and humans, suggesting a zoonotic origin of the infections. The data presented here support a different epidemiology of invasive GCS infections in Portugal, possibly not involving animal exposure, similar to that proposed in a report from the United States (5). In spite of the high phenotypic heterogeneity observed, resulting in a diversity of profiles with the API 20 Strep system, all dubious identifications were confirmed using 16S rRNA sequencing as the "gold standard," substantiating previous suggestions of the usefulness of this system for species identification of GGS and GCS (26, 27).

More than 150 *emm* types are currently recognized for GAS, and a smaller number were found exclusively in GGS or GCS (http://www.cdc.gov/ncidod/biotech/strep/strepindex.htm). We identified 22 distinct *emm* types, including 3 novel types, among the 116 isolates tested; however, none of the *emm* types identified had also been found in GAS, in spite of previous reports documenting the occurrence of typical GAS *emm* types among GGS (21).

PFGE analysis revealed considerable genetic diversity among SDE isolates, indicating that there are several distinct clones causing human infection. Nevertheless, the prevalences of the different clones among this streptococcal population differed considerably, and three large clones accounted for over half of the isolates. Since none of the clones appeared to result from a specific outbreak, it is conceivable that some particular characteristics in the genomes of these strains are responsible for their predominance.

Isolates bearing either C or G group antigen were not clearly separated from each other by PFGE, which is in agreement with the classification of these isolates into a single taxon and confirms previous data based on whole-cell protein profiles (25). An

interesting observation was that PFGE-defined clones did not include a single *emm* type but, in contrast to GAS (10), predominantly comprised isolates with distinct *emm* types. Many isolates with diverse *emm* types clustered with higher Dice coefficients than those presenting the same *emm* type, suggesting that it is unlikely that the observed clustering of isolates with distinct *emm* types is due to a lack of discriminatory power of PFGE. This observation supports previous suggestions of higher rates of genetic exchange among GGS and GCS than among GAS (18). The only exception to the heterogeneity of genetic backgrounds associated with each *emm* type was the most prevalent type in our study, *stG10*, which was found associated with the largest clone (D_{33}) only.

Numerous studies have identified a number of host factors, including malignancy, predisposing to invasive GGS or GCS infection (6). Studies on bacterial factors promoting invasiveness have focused on STSS and its association with the expression of superantigens (16, 19). Although none of the isolates in our collection was associated with a diagnosis of STSS, 24% (n = 28) of our isolates were from invasive disease. These isolates were recovered from normally sterile sites of patients that were hospitalized at the time of isolation, attesting to the severity of the infections and the aggressiveness of the strains. We investigated if either individual *emm* types or clones were more associated with invasive disease than would be expected by chance by calculating ORs comparing the prevalences of individual *emm* types and clones in invasive and noninvasive infections. While no clone was significantly associated with invasiveness, two *emm* types (*stG10* and *stG2078*) had significant ORs.

The presence of *stG10* in a single clone precludes the evaluation of whether the overrepresentation among invasive isolates is due to this *emm* type or to the clonal expansion of a specific strain bearing this *emm* type. However, the nonrandom distribution of invasive isolates within this clone points to the existence of a specific genetic background responsible for the increased virulence, where factors other than the M protein may also play a role.

In contrast to *stG10, stG2078* isolates had a polyclonal origin. Therefore, the observed overrepresentation among invasive isolates cannot be attributable to clonal expansion of a given strain. In this case, our observations suggest that isolates with *stG2078* have a higher probability of causing invasive disease independently of their genetic background, suggesting a role for the product of the *emm* gene or that of a closely linked gene. Heterogeneity in the genomic region flanking the *emm* gene has

already been described among GGS and GCS (14). Although this variability was not correlated with invasiveness, a number of virulence-related genes are known to cluster in this region in GAS, and it is conceivable that alterations of the same region in GGS and GCS lead to changes in virulence. A further indication that strains with *stG2078* have increased invasive potential is the fact that this *emm* type was not more frequently detected globally than other *emm* types. Interestingly, *stG2078* has previously been reported as one of the dominant *emm* types (present in 3 of 21 isolates) among isolates causing invasive disease (17) and was the only *emm* type expressed by two isolates among a collection of GGS and GCS (*n*= 12) associated with STSS (16). In contrast, a study conducted in Israel failed to identify any *stG2078* isolate among 94 GGS from cases of bacteremia (6).

Although the potential of isolates with specific M proteins to cause certain disease manifestations remains to be established, numerous studies point to a nonrandom M type distribution among invasive isolates of GAS (7). Our observations suggest that specific *emm* types are associated with increased invasiveness of GGS and GCS strains. Nevertheless, host factors as well as bacterial traits other than the M protein may contribute to the frequency of invasive disease. This is supported by the finding of *emm* type variability among invasive isolates (10 *emm* types among 22 invasive isolates) reported in this paper and elsewhere (6, 16). Geographic differences in the prevalences of particular clones or *emm* types may also play an important role, since a study of GGS bacteremia in Israel failed to identify any *stG2078* isolates and type *stG10* was represented by a small number of isolates (3 of 94) (6). This study also identified *stG6792* (*n* = 3) and *stG166b* (*n* = 6) among isolates causing bacteremia, whereas in our collection neither of these *emm* types was identified in invasive isolates. However, the Israeli study does not provide information on the frequency of these *emm* types in noninvasive disease, precluding a direct comparison, but these differences highlight the importance of conducting studies similar to that presented here in diverse geographic locations.

The data presented highlight the importance of studying both GGS and GCS to fully evaluate the prevalence and severity of infections by SDE and suggest an association between *emm* type and invasive disease potential in this bacterium.

ACKNOWLEDGMENTS

This work was partly supported by Fundação Calouste Gulbenkian, Portugal.

Members of the Portuguese Group for the Study of Streptococcal Infections are as follows: Luís Lito and Maria José Salgado, Hospital de Santa Maria, Lisbon; Filomena Martins and Maria Ana Pessanha, Hospital de São Francisco Xavier, Lisbon; José Diogo, Ana Rodrigues, and Isabel Nascimento, Hospital Garcia de Orta, Almada; Fernanda Cotta and J. Correia da Fonseca, Hospital de São João, Porto; Paulo Lopes, Ismália Calheiros, Luísa Felício, Cacilda Magalhães, and Lourdes Sobral, Centro Hospitalar de Vila Nova de Gaia; Valquíria Alves and Antónia Read, Hospital Pedro Hispano, Matosinhos; Ana Paula Castro, Hospital de Vila Real; and Ana Paula M. Vieira and Francisco B. Moniz, Hospital Senhora da Oliveira, Guimarães.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: M. D. Pinho, J. Melo-Cristino and M. Ramirez. Performed the experiments: M. D. Pinho. Analyzed the data: M. D. Pinho, J. Melo-Cristino and M. Ramirez. Contributed reagents/materials/analysis tools: J. Melo-Cristino and M. Ramirez. Wrote the paper: M. D. Pinho, J. Melo-Cristino and M. Ramirez **.**

REFERENCES

1. **Beall B, Facklam RR, Elliott JA, Franklin AR, Hoenes T, Jackson D, Laclaire L, Thompson T, Viswanathan R**. 1998. Streptococcal *emm* types associated with T-agglutination types and the use of conserved *emm* gene restriction fragment patterns for subtyping group A streptococci. J. Med. Microbiol. **47**:893-898.

2. **Bert F, Branger C, Lambert-Zechovsky N**. 1997. Pulsed-field gel electrophoresis is more discriminating than multilocus enzyme electrophoresis and random amplified polymorphic DNA analysis for typing pyogenic streptococci. Curr. Microbiol. **34**:226-229.

3. **Bradley SF, Gordon JJ, Baumgartner DD, Marasco WA, Kauffman CA**. 1991. Group C streptococcal bacteremia: analysis of 88 cases. Rev. Infect. Dis. **13**:270-280.

4. **Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG**. 2003. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. J. Infect. Dis. **187**:1424-1432.

5. **Carmeli Y, Ruoff KL**. 1995. Report of cases of and taxonomic considerations for large-colony-forming Lancefield group C streptococcal bacteremia. J. Clin. Microbiol. **33**:2114-2117.

6. **Cohen-Poradosu R, Jaffe J, Lavi D, Grisariu-Greenzaid S, Nir-Paz R, Valinsky L, Dan-Goor M, Block C, Beall B, Moses AE**. 2004. Group G streptococcal bacteremia in Jerusalem. Emerg. Infect. Dis. **10**:1455-1460.

7. **Cunningham MW.** 2000. Pathogenesis of group A streptococcal infections. Clin. Microbiol. Rev. **13**:470-511.

8. **Daniel WW.** 1999. Biostatistics: a foundation for analysis in the health sciences, 7th ed. Wiley, New York, N.Y.

9. **Efstratiou A.** 1997. Pyogenic streptococci of Lancefield groups C and G as pathogens in man. Soc. Appl. Bacteriol. Symp. Ser. **26**:72S-79S.

10. **Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE**. 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between *emm* type and clone. Infect. Immun. **69**:2416-2427.

11. **Eriksson BK, Norgren M, McGregor K, Spratt BG, Normark BH**. 2003. Group A streptococcal infections in Sweden: a comparative study of invasive and noninvasive infections and analysis of dominant T28 *emm28* isolates. Clin. Infect. Dis. **37**:1189-1193.

12. **Facklam R**. 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin. Microbiol. Rev. **15**:613-630.

13. **Figueira-Coelho J**, **Ramirez M, Salgado MJ, Melo-Cristino J**. 2004. *Streptococcus agalactiae* in a large Portuguese teaching hospital: antimicrobial susceptibility, serotype distribution, and clonal analysis of macrolide-resistant isolates. Microb. Drug Resist. **10**:31-36.

14. **Geyer A, Schmidt KH**. 2000. Genetic organisation of the M protein region in human isolates of group C and G streptococci: two types of multigene regulator-like (*mgrC*) regions. Mol. Gen. Genet. **262**:965-976.

15. **Haidan A, Talay SR, Rohde M, Sriprakash KS,Currie BJ, Chhatwal GS**. 2000. Pharyngeal carriage of group C and group G streptococci and acute rheumatic fever in an Aboriginal population. Lancet **356**:1167-1169.

16. **Hashikawa S, Iinuma Y, Furushita M, Ohkura T, Nada T, Torii K, Hasegawa T, Ohta M**. 2004. Characterization of group C and G streptococcal strains that cause streptococcal toxic shock syndrome. J. Clin. Microbiol. **42**:186-192.

17. **Igwe EI, Shewmaker PL, Facklam RR, Farley MM, van Beneden C, Beall B.** 2003. Identification of superantigen genes *speM*, *ssa*, and *smeZ* in invasive strains of beta-hemolytic group C and G streptococci recovered from humans. FEMS Microbiol. Lett. **229**:259-264.

18. **KaliA A, Enright MC, Spratt BG, Bessen DE.** 2001. Directional gene movement from human-pathogenic to commensal-like streptococci. Infect. Immun. **69**:4858-4869.

19. **Korman TM, Boers A, Gooding TM, Curtis N, Visvanathan K.** 2004. Fatal case of toxic shock-like syndrome due to group C streptococcus associated with superantigen exotoxin. J. Clin. Microbiol. **42**:2866-2869.

20. **McGregor KF, Spratt BG, Kalia A, Bennett A, Bilek N, Beall B, Bessen DE.** 2004. Multilocus sequence typing of *Streptococcus pyogenes* representing most known *emm* types and distinctions among subpopulation genetic structures. J. Bacteriol. **186**:4285-4294.

21. **Schnitzler N, Podbielski A, Baumgarten G, Mignon M, Kaufhold A**. 1995. M or M-like protein gene polymorphisms in human group G streptococci. J. Clin. Microbiol. **33**:356-363.

22. **Serrano I, Melo-Cristino J, Carriço JA, Ramirez M**. 2005. Characterization of the genetic lineages responsible for pneumococcal invasive disease in Portugal. J. Clin. Microbiol. **43**:1706-1715.

23. **Sylvetsky N, Raveh D, Schlesinger Y, Rudensky B, Yinnon AM.** 2002. Bacteremia due to beta-hemolytic *Streptococcus* group G: increasing incidence and clinical characteristics of patients. Am. J. Med. **112**:622-626.

24. **Takahashi T, Kaneko M, Mori Y, Tsuji M, Kikuchi N, Hiramune T.** 1997. Phylogenetic analyses of *Staphylococcus* based on the 16S rDNA sequence and assignment of clinical isolates from animals. J. Vet. Med. Sci. **59**:775-783.

25. **Vandamme P, Pot B, Falsen E, Kersters K, Devriese LA**. 1996. Taxonomic study of Lancefield streptococcal groups C, G, and L (*Streptococcus dysgalactiae*) and proposal of *S. dysgalactiae* subsp. *equisimilis* subsp. nov. Int. J. Syst. Bacteriol. **46**:774-781.

26. **Vieira VV, Teixeira LM, Zahner V, Momen H, Facklam RR, Steigerwalt AG, Brenner DJ, Castro AC**. 1998. Genetic relationships among the different phenotypes of *Streptococcus dysgalactiae* strains. Int. J. Syst. Bacteriol. **48**:1231-1243.

27. **Woo PC, Fung AM, Lau SK, Wong SS, Yuen KY**. 2001. Group G beta-hemolytic streptococcal bacteremia characterized by 16S ribosomal RNA gene sequencing. J. Clin. Microbiol. **39**:3147-3155.

28. **Yuen KY, Seto WH, Choi CH, Ng W, Ho SW, Chau PY**. 1990. *Streptococcus zooepidemicus* (Lancefield group C) septicaemia in Hong Kong. J. Infect. **21**:241-250.

29. **Zaoutis T, Attia M, Gross R, Klein J**. 2004. The role of group C and group G streptococci in acute pharyngitis in children. Clin. Microbiol. Infect. **10**:37-40.

CHAPTER 3

Population genetics of *Streptococcus dysgalactiae* subsp. *equisimilis* reveals widely dispersed clones and extensive recombination

The contents of this chapter were presented in:

McMillan DJ, Bessen DE, Pinho M, Ford C, Hall JS, Melo-Cristino J, Ramirez M

2010. PLoS One 23; **5**:e1174

ABSTRACT

Streptococcus dysgalactiae subsp. *equisimilis* (SDE) is an emerging global pathogen that can colonize and infect humans. Although most SDE isolates possess the Lancefield group G carbohydrate, a significant minority have the group C carbohydrate. Isolates are further sub-typed on the basis of differences within the *emm* gene. To gain a better understanding of their molecular epidemiology and evolutionary relationships, multilocus sequence typing (MLST) analysis was performed on SDE isolates collected from Australia, Europe and North America. The 178 SDE isolates, representing 37 *emm* types, segregate into 80 distinct sequence types (STs) that form 17 clonal complexes (CCs). Eight STs recovered from all three continents account for more than 50% of the isolates. Thus, a small number of STs are highly prevalent and have a wide geographic distribution. Both ST and CC strongly correlate with group carbohydrate. In contrast, eleven STs were associated with more than 1 *emm* type, suggestive of recombinational replacements involving the *emm* gene; furthermore, 35% of the *emm* types are associated with genetically distant STs. Data also reveal a history of extensive inter- and intra-species recombination involving the housekeeping genes used for MLST. Sequence analysis of single-locus-variants identified through goeBURST indicates that genetic change mediated by recombination occurred approximately 4.4 times more frequently than by point mutation. A few genetic lineages with an intercontinental distribution dominate among SDE causing infections in humans. The distinction between group C and G isolates reflects recent evolution, and no long-term genetic isolation between them was found. Lateral gene transfer and recombination involving housekeeping genes and the *emm* gene are important mechanisms driving genetic variability in the SDE population.

INTRODUCTION

Most streptococci displaying beta-hemolysis fall within the pyogenic branch of the 16S rRNA-based taxonomy, and are pathogens or commensals of mammalian hosts (1, 2). Two species within the pyogenic branch - *Streptococcus dysgalactiae* subsp. *equisimilis* (SDE) and *Streptococcus pyogenes* (group A streptococcus, GAS) - colonize and/or infect the respiratory tract and skin of the human host (2). Whereas GAS is an important human pathogen, SDE is largely considered to be a commensal organism. However, numerous studies report that SDE can cause disease among otherwise healthy individuals (2–5). The disease spectrum of SDE infection is similar to that of GAS, and includes pharyngitis, post-streptococcal glomerulonephritis, cellulitis, necrotizing fasciitis, septicemia, and streptococcal toxic shock syndrome (6–9). Furthermore, in some geographic regions where streptococcal diseases are endemic, surveillance studies report higher rates of throat colonization by SDE than by GAS (10, 11).

The surface-exposed Lancefield group carbohydrate is an important cell wall antigen that aids in distinguishing between several of the beta-hemolytic streptococcal species. *S. pyogenes* almost exclusively expresses the group A carbohydrate (1). Although the vast majority of SDE isolates have group G carbohydrate, and are often referred to as group G streptococci, a significant minority of SDE isolates have group C carbohydrate; very rarely do SDE harbor the group A or L carbohydrate (1, 12). Among the GAS and SDE populations, differences in the sequences of individual *emm* genes are widely used for intra-species strain typing. At present, more than 200 GAS and approximately 50 SDE *emm* types are recognized (http://www.cdc.gov/ncidod/biotech/strep/strepblast.htm). Although one report finds SDE isolates expressing *emm* types *stG2078* or *stG10* to have enhanced invasive disease potential (5), most studies have failed to uncover disease associations among SDE *emm* types (13–15). In contrast, associations between *emm* type and specific diseases are well established for GAS (16–19).

Comparative genome hybridization studies using a microarray containing probes corresponding to genes encoding virulence factors and putative surface proteins failed to reveal clear cut associations between *emm* type and gene content among SDE (13). Multilocus sequence typing (MLST) is a nucleotide sequence-based method that uses core housekeeping genes to characterize genetic relationships between isolates of the same species. MLST has been used extensively to study the beta-hemolytic GAS (20–

22) and *S. agalactiae* (23–24) populations, and was recently used to investigate genetic relationships amongst 61 geographically restricted SDE isolates (25). In the present study, an intercontinental collection of SDE isolates is characterized by MLST and *emm* typing, and the geographic distribution of the identified clones and their genetic relationships are defined.

MATERIALS AND METHODS

Bacterial Strains

The 117 SDE isolates analyzed in this study were collected from Australia (*n* = 55), Portugal (*n* = 36) and the United States (*n* = 11). Another 15 SDE isolates were obtained from other countries or had no associated geographic information (Table S3.1). All isolates were classified as SDE on the basis of isolation from a human host, beta-hemolysis following growth on sheep blood-containing agar, the presence of group C or G carbohydrate, and the presence of a recognized *emm* type. *emm* type was determined by nucleotide sequence typing as described (38), and *emm* type was assigned using the BLASTn-*emm* server (http://www.cdc.gov/ncidod/biotech/strep/strepblast.htm). Group carbohydrate was determined using the latex bead agglutination test. Forty-five isolates were obtained from normally sterile tissue or fluid and classified as causing invasive disease; 59 isolates were recovered from other non-sterile sites and classified as causing noninvasive infections. Information on the 61 invasive isolates recovered from the United States, and reported in a previous study (25), was integrated into the current analyses to obtain a global overview of the relationships between GGS and GCS causing infections in humans.

MLST

Seven housekeeping genes are used for MLST of SDE. With the exception of *atoB*, the housekeeping genes used for MLST of SDE are the same as those used in the GAS MLST scheme (38). The seventh allele, *atoB*, (also called *yqi*Z) has been described (25); the allele name has been changed for this report, in order to reduce confusion with the GAS MLST gene, *yqiL*, which occupies a different *locus*. Primer sequences used for PCR amplification are listed in Table S3.2. Although our MLST scheme was developed independently from that reported previously (25), the data is directly comparable because the same genes, and regions within genes, are targeted for sequencing. All PCR products were sequenced in both the forward and reverse directions; newly identified alleles and alleles defining new STs were re-sequenced in order to validate the initial findings. Unique alleles at each *locus* were assigned individual allele numbers. The combination of the seven allele numbers for each isolate was used to define the sequence type (ST). MLST data for SDE is available at http://sdse.mlst.net/.

Data analysis

goeBURST (http://goeburst.phyloviz.net/) (39), which uses the same clustering rules as eBURST (40) but provides a global optimal solution, was used to determine the relationships between STs. Clonal complexes (CCs) are defined as STs that are linked through single-locus-variants (SLVs) and are named on the basis of the predicted founder ST, which is the ST having the most SLVs. In cases where a CC contains only two STs, the lower numbered ST was used to define the CC. Isolates that share four of seven alleles (i.e. triple-locus-variants, TLVs) were used to define larger, more distantly related clonal groups.

The Simpson Index of Diversity (SID) and Wallace coefficients (W) were calculated as described previously (41) using www.comparingpartitions.info. A SID value equal to one signifies that the typing method distinguishes between all isolates, whereas a SID value equal to zero means that all isolates are identical. The W coefficient provides a finer comparison between two typing methods, since the value indicates the probability that two strains classified as the same type by one method are also classified as the same type by the other method. A high value of the W coefficient (values close to 1) indicates that partitions defined by a given method could have been predicted from the results of another method, suggesting that the use of both methodologies could be redundant. Nucleotide diversity (π), nonsynonymous (d_n) and synonymous substitution rates (d_s) were calculated using DnaSP (version 5) (42).

Phylogenetic analysis

Phylogenetic relationships amongst individual housekeeping alleles was examined using the neighbor joining (NJ) method with Jukes-Cantor substitution algorithm model as implemented in MEGA4 (43); bootstrapping (1000 replicates) was used to ascertain support for branches. For concatenated housekeeping alleles, evolutionary history was inferred using the NJ method or maximum parsimony method (PAUP 4.0).

Recombination and point mutation

Empirical estimates of the number of mutation and recombination events contributing to the diversification of SLVs were made as previously described (21, 44), with additional modifications. The nucleotide sequence differences in the mismatched allele among the pair of STs that define an SLV are scored as genetic changes that arise

due to either mutation or recombination. Nucleotide differences between the two variant alleles at greater than 1 nucleotide site is scored as a probable recombination event. Single nucleotide differences between the two variant alleles that do not occur in other any other alleles among the SDE set of strains are scored as a change likely due to point mutation. If the single nucleotide polymorphism is present in two or more alleles assigned to different CCs, the genetic change is scored as likely due to recombination. The relative ratio of recombination events versus mutation events was then determined. To calculate the per site recombination/mutation (r/m) ratio, the total number of nucleotide sites that change due to recombination were divided by the total number of nucleotide sites that change due to mutation.

Recombination among SDE and between the SDE and GAS at each individual gene was evaluated using Recombination Detection Program (RDP) (26), which implements a large number of methods (RDP, GENECONV, MaxChi, Chimaera, SiScan, 3Seq) for detecting intragenic recombination. For these analyses the entire collection of GAS alleles was downloaded from the MLST database (http://spyogenes.mlst.net). Since a high number of comparisons were performed, the p values reported are corrected for multiple tests. SplitsTrees4 was also used to assess intragenic recombination (45), excluding parsimony uninformative sites, and using the neighbor net algorithm, uncorrected P distance, and the Phi test for recombination (46).

RESULTS

Molecular typing

MLST was used to characterize 117 SDE isolates collected from three continents (Table S3.1). Most of the isolates selected for MLST were derived from large independent collections and were chosen, in part, based on prior knowledge of their *emm* type and geographic site of isolation, with the goal being the compilation of a genetically diverse data set.

The isolates represent 24 of the approximately 50 known SDE *emm* types, and include strains bearing the group G or group C carbohydrate. A summary of the epidemiological features of the isolates is provided in Table 3.1.

With the inclusion of previously published MLST data on 61 invasive SDE isolates obtained from the USA (25), a total of 37 *emm* types are represented among a larger set of 178 isolates, which is used for the analyses in this report. The frequencies of the different *emm* types span a wide range; however, only 38% of the *emm* types account for the majority (74%) of the isolates.

Among the 178 SDE isolates, the number of alleles identified for each housekeeping gene *locus* ranges from 10 for *gtr*, *murI* and *mutS*, to 22 for *xpt* (Table 3.2). The *gki*, *recP*, and *xpt loci* exhibit the highest level of nucleotide sequence diversity (π). The percentage of polymorphic nucleotide sites ranges from 2.7 (n = 12) for *murI*, to 10.1 (n = 50) for *gki*. A significant portion of the polymorphism observed in *gki* can be attributed to *gki12*, a highly divergent allele with greater similarity to GAS *gki* alleles than to SDE alleles (Figure S3.1); when the divergent *gki12* allele is removed from the analysis, the nucleotide diversity of *gki* falls from 0.021 to 0.012, and the percentage of polymorphic sites drops from 10.1 to 5.4 (n = 27). The d_n/d_s ratio is less than one for each of the seven housekeeping genes, consistent with stabilizing selection.

Relationships among STs

The seven housekeeping alleles of each of the 178 isolates yield a total of 80 distinct allelic profiles, referred to as sequence types (STs). Of the 80 STs, 37 are newly identified in this study and 43 were previously defined by Ahmad *et al.* (25). A minority of STs (10%) account for a disproportionate number (approximately 50%) of the total

TABLE 3.1. Characteristics of SDE isolates included in this study

Collection site	No. of isolates	No. off different partitions (SID ± 95% CI) [a]		STs unique to collection site[b]	No. of isolates in Lancefield group		No. of invasive/ non-invasive isolates
		emm types	STs		C	G	
Australia	55	17 (0.926 ± 0.030)	23 (0.937 ± 0.033)	14	47	8	24/25
Portugal	36	17 (0.951 ± 0.021)	22 (0.956 ± 0.033)	11	28	8	10/26
USA	72	34 (0.984 ± 0.008)	45 (0.975 ± 0.014)	33	48	23	70/1
NYMC	11	8 (0.927 ± 0.094)	10 (0.981 ± 0.045)	5	6	5	9/0
CDC	61	33 (0.983 ± 0.005)	37 (0.972 ± 0.016)	28	42	18	61/1
Other	15	12 (0.971 ± 0.030)	13 (0.981 ± 0.030)	10	7	7	1/6
Total	178	37 (0.961 ± 0.006)	80 (0.966 ± 0.011)	na	131	46	106/59

[a] SID – Simpsons Index of diversity; CI – Confidence interval
[b] na – not applicable

TABLE 3.2. Housekeeping genes used for MLST of SDE

Gene	ORF[a]	Size of partial gene	No. of alleles	No. of variant positions		π	d_n	d_s	d_n/d_s
				Nucleotide (%)	Amino acid				
Glucose kinase (*gki*)	SDEG_1515	498	12	50 (10.1)	7	0.021	0.0035	0.0736	0.047
Glutamine transport protein (*gtr*)	SDEG_1494	450	10	15 (3.3)	7	0.010	0.0059	0.0248	0.240
Glutamate racemase *(murI*)	SDEG_0413	438	10	12 (2.7)	1	0.012	0.0068	0.0142	0.479
DNA mismatch repair protein (*mutS*)	SDEG_2091	405	10	27 (6.7)	7	0.016	0.0062	0.0480	0.129
Transketolase (*recP*)	SDEG_1735	459	20	37 (8.1)	3	0.034	0.0008	0.1472	0.006
Xanthine phosphoribosyl transferase (*xpt*)	SDEG_0895	450	22	38 (8.4)	10	0.021	0.0049	0.0736	0.067
Acetoacetyl-coathioloase (*atoB*)	SDEG_1700	434	12	18 (4.1)	5	0.011	0.0035	0.0330	0.106

[a] Based on ORF number in the GGS_124 genome (GenBank number AP010935)

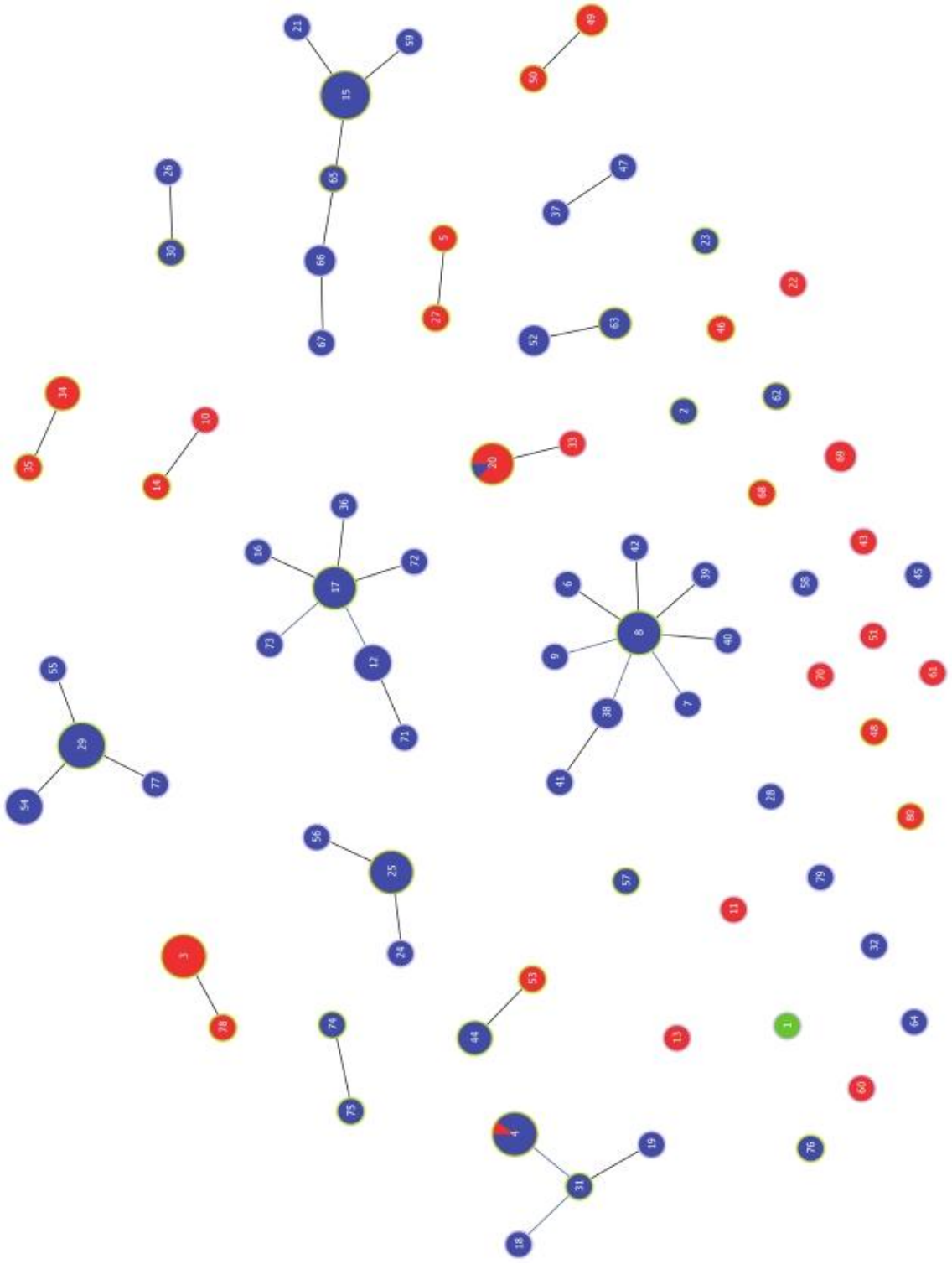

FIGURE 3.1. goeBURST diagram of relationships between 178 global SDE isolates

The size of each circle is proportional to the number of isolates with that particular ST in a logarithmic scale. STs assigned to the same CC are linked by straight lines. Blue circles represent isolates that have the group G carbohydrate. Red circles represent isolates expressing the group C carbohydrate. Whenever isolates of the same ST have different group carbohydrates, the number of isolates bearing the same carbohydrate is proportional to the respective color. The green circle represents the single isolate expressing the group L carbohydrate.

SDE isolates under evaluation. The vast majority of STs (62, or 77%) are represented by only one SDE isolate. The most prevalent ST (ST15) is represented by 20 isolates. Eight STs, each represented by eight or more isolates, account for 51% of the 178 SDE isolates characterized by MLST.

The 80 STs can be grouped into 17 clonal complexes (CCs) by goeBURST, whereby the connected STs are single-locus-variants (SLVs) of at least one other ST in the group, differing at only one housekeeping gene (Figure 3.1). However, only six of the 17 CCs contain more than two STs. CC8 contains the highest number of STs (n = 9), whereas CC15 (n = 26) and CC29 (n = 20) contain the most isolates.

Twenty-six STs, representing 15% of the 178 isolates, are singletons and differ from all other STs by more than two housekeeping alleles. When clusters are constructed linking STs that are up to triple-locus-variants (TLVs) of each other, 66 STs are grouped into a single major cluster, whereas only three STs remain ungrouped singletons (Figure S3.3), indicating that several intermediate genotypes probably exist, but have not yet been sampled.

Relationships between *emm* type and ST

The overall correspondence between ST and *emm* type, as determined by the Wallace Coefficient, is low (ST vs *emm* type W = 0.473, $CI_{95\%}$ 0.332–0.542; *emm* type vs ST W = 0.384, $CI_{95\%}$ 0.311–0.456) reflecting the fact that most *emm* types are found in multiple STs, and that the same ST can harbor different *emm* types (Table 3.3 and Table 3.4). The correspondence between *emm* type and CC (W = 0.551, $CI_{95\%}$ 0.478 to 0.625) is also weak.

STs associated with multiple *emm* types most likely arose via recombinational replacement of the *emm* gene; they are referred to as *emm* variable STs. Of the 18 STs represented by more than one isolate, 11 STs (61%) are associated with more than one *emm* type (Table 3.3). Five STs are associated with two *emm* types (ST34, 20, 29, 52, 63), two STs (ST17, 25) are associated with three *emm* types, one ST (ST3) is associated with four *emm* types, two STs (ST4, ST8) are associated with five *emm* types and one ST (ST15) is associated with 7 different *emm* types.

TABLE 3.3. Relationship between ST and *emm* type

ST	No. of isolates	Associated *emm* types	No. of *emm* types
15	20	*stC839, stG10, stG166b, stG2078, stG245, stG6, stG652*	7
8	10	*stC839, stG11, stG480, stG643, stG7860*	5
4	10	*stC36, stC5344, stG6792, stG97, stG7882*	5
3	10	*EMM57, stC1400, stC839, stG653*	4
25	9	*stG166b, stG5420, stG6*	3
17	9	*stC74a, stG2078, stG485*	3
63	2	*stG6, stG652*	2
52	2	*stG6, stG643*	2
20	8	*stC6979, stG62647*	2
34	3	*stC1400, stG5063*	2
29	14	*stC74a, stG485*	2

TABLE 3.4. Relationship between *emm* type and ST

***emm* type**	**No. of isolates**	**Associated STs**	**No. of:**			
			STs	**CC_{SLV}[a]**	**CC_{DLV}[b]**	**Distant STs[c]**
stG6	13	15, 24, 25, 44, 52, 58, 62, 63	8	4	2	2
stG480	11	7, 8, 38, 39, 40, 41, 67	7	2	2	2
stC1400	8	3, 28, 34, 46, 64, 66	6	3	1	4
stG485	8	17, 29, 37, 47, 55, 69	6	3	2	2
stG643	9	8, 12, 22, 48, 52, 73	6	3	4	5
stG652	7	15, 32, 59, 61, 63, 71	6	3	2	
stC6979	8	9, 19, 20, 54, 80	5	4	4	3
stC36	6	4, 45, 49, 50, 68	5	2	1	2
stC74a	15	17, 29, 70, 77	4	2	3	3
stC839	7	3, 8, 15, 78	4	3	2	3
stG166b	5	15, 25, 56, 65	4	2	2	2
stG11	5	6, 8, 42	3	1	1	
stG2078	9	15, 17, 72	3	2	2	2
stG245	3	15, 21, 36	3	2	2	2
stG4831	3	74, 75, 76	3	1	0	
stG62647	9	20, 33, 60	3	1	1	
stG6792	6	4, 31, 51	3	1	1	
EMM57	3	3, 57	2	1	0	2
stC5344	3	3, 43	2	1	1	
stC6746	2	5, 27	2	1	0	
stC9431	2	13, 14	2	1	1	
stG5063	2	2, 34	2	1	0	
stG840	2	26, 30	2	1	1	
stG7882	2	4, 18	2	1	1	

[a] CCSLV – Clonal complex based on single-locus-variant relationships.

[b] CCDLV – Clonal complex based on double-locus-variant relationships.

[c] Number of STs sharing the same *emm* type and differing from all other STs harboring that *emm* type at greater than five housekeeping alleles.

The same *emm* type is often found in association with multiple STs (Table 3.4). The association of a given *emm* type with multiple STs can arise following diversification of housekeeping genes, or by horizontal transfer of the *emm* gene. An estimate of the horizontal movement of *emm* is made by enumerating the number of distant STs harboring the same *emm* type, whereby distant STs are defined as having five or more housekeeping allele differences to any other ST that shares the same *emm* type; for a given CC, only one representative ST is counted.

Thirteen *emm* types are associated with distant STs: five *emm* types are associated with more than two distant STs and eight *emm* types are found among a single pair of distant STs, whereas five *emm* types are associated with three or more distant STs (Table 3.4). The most promiscuous *emm* type is *stG643*, found among five genetically distant strains. A total of 21 horizontal transfer events involving *emm* genes are evident in the SDE data set. Taken together, the data provides strong support for the hypothesis that *emm* genes of SDE undergo extensive lateral exchange between strains.

Relationships between group carbohydrate and ST

The group specific carbohydrate of the streptococcal cell wall can be used to discriminate among β-hemolytic streptococcal species. The majority of SDE isolates (74%) in this study express the group G carbohydrate. A sharp distinction between STs associated with strains expressing the group G versus C carbohydrate is observed by goeBURST (Figure 3.1). Isolates representing 54 STs have group G carbohydrate (group G streptococci, GGS), whereas 27 STs are associated with strains expressing group C carbohydrate (group C streptococci; GCS). A single isolate has the group L carbohydrate.

Only two STs (ST4 and ST20) have isolates associated with both group C and G carbohydrates; for ST4, nine of 10 isolates are GGS and for ST20, seven of eight isolates are GCS. Overall, ST and group carbohydrate, whose biosynthesis *locus* is unknown, display very strong linkage (W = 0.970, $CI_{95\%}$ 0.939 to 1.000). The correspondence between CCs and group carbohydrate is also high, with nine CCs containing only GGS isolates and six CCs restricted to GCS isolates (W = 0.979, $CI_{95\%}$ 0.957 to 1.000), indicating that STs belonging to the same genetic lineage almost always also share the same group carbohydrate.

TABLE 3.5. Distribution of housekeeping alleles among GCS and GGS isolates

Housekeeping gene *locus*[a]	Percentage of alleles:		
	Shared by GCS and GGS	**Restricted to GCS**	**Restricted to GGS**
murI	30	30	40
xpt	32	36	32
gtr[b]	56	33	11
gki	33	25	43
atoB	50	33	17
recP	35	35	30
mutS	40	30	30
Total for all alleles	38	36	29

[a] Presented in order of the *locus* position on the genome of strain GGS_124.
[b] Excludes *gtr06* which is restricted to group L.

For each of the seven housekeeping *loci*, the relative distribution of alleles among GCS and GGS isolates was evaluated (Table 3.5). Overall, 38% of the housekeeping alleles are shared among GCS and GGS isolates. Approximately 36% and 29% of the housekeeping alleles are restricted to GCS and GGS isolates, respectively. This finding shows that there is a common housekeeping gene pool that is shared among numerous GCS and GGS isolates, despite the highly restricted associations that are observed between group carbohydrate and ST.

Relationships between group carbohydrate and *emm* type

Unlike ST, which is largely restricted to a single group carbohydrate form, 13 (35%) of the 37 *emm* types are found in association with both GGS and GCS isolates (W = 0.821, $CI_{95\%}$ 0.764 to 0.876). Eight of the variable associations between *emm* type and group carbohydrate likely arose following horizontal transfer of an *emm* gene to a strain having a different group carbohydrate (data not shown). However, it remains possible that lateral movement of genes encoding carbohydrate biosynthetic enzymes also contributes to the generation of diversity among SDE, although the frequency of this event is probably low because carbohydrate-variable STs are rare.

Geographic distribution of genetically diverse SDE isolates

To examine the global distribution of SDE clones, the ST and *emm* type of the isolates recovered from Australia, Europe and North America were compared. Clonal diversity based on either ST or *emm* type, as measured by the Simpson Index of

Diversity, was high (>0.9) for isolates collected from each of the three primary locations (Table 3.1).

The eight STs that were recovered from all three continents also represent the STs with the highest overall prevalence (ST3, 4, 8, 15, 17, 20, 25, 29) (Table S3.1; Figure 3.2). Furthermore, five of the eight highly prevalent STs are predicted to be founders of a CC, wherein the founder ST is defined as that having the highest number of SLVs. The data suggest a wide geographical dispersion of founder STs. Of the next 10 most highly prevalent STs (each of which is represented by two to four isolates), four were associated with isolates from two continents (Table S3.1). Fourteen *emm* types were recovered from each of the three continents; 13 of these are represented by the most highly prevalent strains, having greater than 5 SDE isolates per *emm* type. Together the results demonstrate that the most highly prevalent strains of SDE, whereby strain is defined by either ST or *emm* type, are widely disseminated.

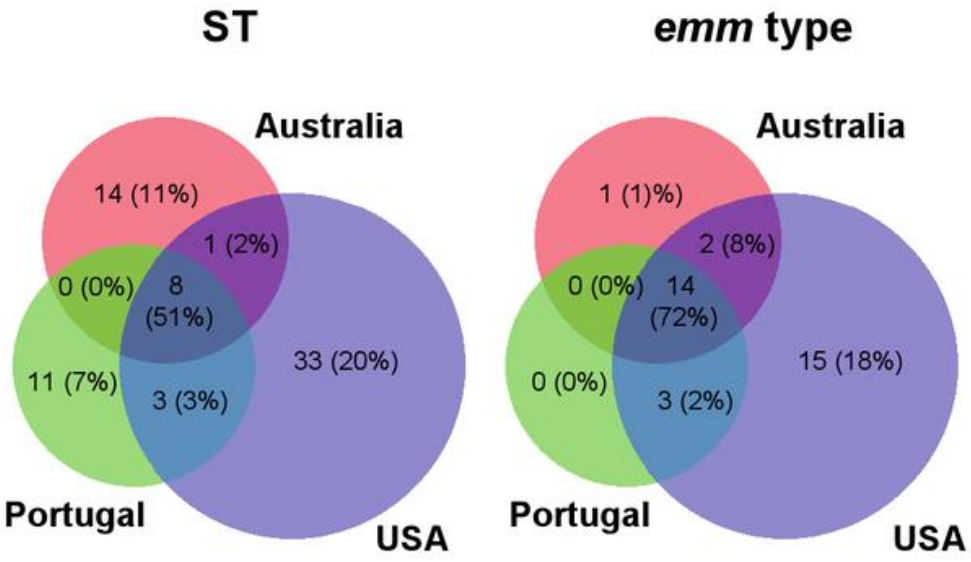

FIGURE 3.2. Venn diagram depicting the distribution of ST and *emm* type across three continents

Unbracketed numbers represent the total number of STs or *emm* types. The numbers in brackets indicate the percentage of total isolates in the entire collection.

Among the 178 SDE isolates, 107 unique combinations of ST, *emm* type, and group carbohydrate were observed. Isolates corresponding to eight of the 107 unique *emm* ST-carbohydrate profiles are widespread and were recovered from all three continents (Table 3.6); another seven were isolated from two continents. Three intercontinental

clones belong to CC8, two of which likely arose from a common ancestor by either genetic diversification at a housekeeping gene *locus* and/or by horizontal transfer of the *emm* gene. The two intercontinental CC17 clones are SLVs, whereas the CC15 and CC25 sets of intercontinental strains arose via lateral exchange of *emm* type. The genetic changes are likely to be ancient events that preceded the intercontinental migration of the founders.

TABLE 3.6. Intercontinental clones of SDE

CC	ST	*emm* type	Group carbohydrate	Australia	Europe	North America
3	3	*stC839*	C	x	x	x
4	4	*stG6792*	G		x	x
8	8	*stG480*	G	x	x	x
8	8	*stG11*	G	x		x
8	38	*stG480*	G	x	x	x
15	15	*stG10*	G	x	x	x
15	15	*stG652*	G	x		x
15	15	*stG166b*	G		x	x
17	17	*stG2078*	G	x	x	x
17	12	*stG643*	G	x		x
20	20	*stG62647*	C	x	x	x
25	25	*stG5420*	G	x	x	x
25	25	*stG6*	G	x	x	
29	29	*stC74a*	G	x	x	x
49	49	*stC36*	C		x	x

Phylogenetic analysis of housekeeping genes

Clonal relationships established via goeBURST are based on the character state of the housekeeping gene allele, and do not take into account the degree of nucleotide sequence heterogeneity. In order to further investigate the relatedness of the housekeeping gene alleles at each *locus*, phylogenetic trees for each gene were constructed by the neighbor joining method. With the exception of *atoB*, these trees included the alleles from *loci* of GAS having the highest percentage nucleotide sequence identity based on BLASTn. Additionally, the housekeeping genes of GAS and SDE share synteny (Table 3.2) (20). In agreement with a previous report (25), several SDE alleles are more similar to GAS alleles than to other SDE alleles (Figures S3.1, S3.2). Both *gki12* and *mutS3* form a cluster with GAS alleles, whereas all *gtr* and *murI* alleles from SDE and GAS segregate into distinct species-specific clusters. The

relationship between *recP* and *xpt* alleles in the two species is more complex, and the phylogenies for the *xpt* and *recP* alleles do not resolve into species-specific clusters.

The seven housekeeping alleles were concatenated (3,134 nucleotide sites) for each of the 80 STs of SDE, and the concatenates used to construct a phylogenetic tree by the maximum parsimony method (Figure 3.3). The relative distribution of STs along the tree branches is highly concordant with the CCs generated via the goeBURST clustering algorithm that used allele character states (Figure 3.1). A striking feature of the phylogenetic tree is that GCS and GGS taxa are highly interspersed and fail to form discrete evolutionary lineages, even in portions of the tree having strong bootstrap support. However, the homoplasy index for the phylogenetic tree is high (0.7331, excluding uninformative characters) and strong bootstrap support is absent for many of the deeper branches. Thus, the phylogeny may be less accurate for long term evolutionary events due to a past history of extensive recombination, but nevertheless, it appears to recapitulate the short term evolution detected by goeBURST. The analysis also provides additional support for the horizontal transfer of housekeeping genes between GCS and GGS organisms.

Phylogenetic trees constructed by the minimum evolution (data not shown) or neighbor joining (Figure S3.4) methods also exhibit high concordance with the CCs generated via goeBURST. The only exception is CC15, which is divided into two or three small subclusters. However, like the maximum parsimony tree (Figure 3.3), there is little bootstrap support for deep branches in the neighbor joining tree (Figure S3.4).

Role of recombination in genetic change in SDE

An analysis of recombination events between the GAS and SDE housekeeping gene alleles, using the RDP suite (26), predicts extensive recombination between GAS and SDE in the *gki* gene. The *gki75* and *gki102* alleles of GAS were identified as having derived from recombination with SDE alleles ($p<0.00001$ and $p = 0.00037$, respectively), whereas the *gki12* and *gki4* of SDE appear to result from recombination with GAS alleles (both with $p<0.00001$). These data account for the high nucleotide percentage diversity observed in the *gki locus*. Recombination between GAS and SDE could also be identified using RDP in the *recP* gene. The GAS *recP21*, *recP40*, *recP54*, *recP71*, *recP85* are presumed to have resulted from recombination with SDE alleles ($p = 0.02469$, $p = 0.01073$, $p = 0.00050$, $p = 0.00900$ and $p = 0.00900$, respectively). In

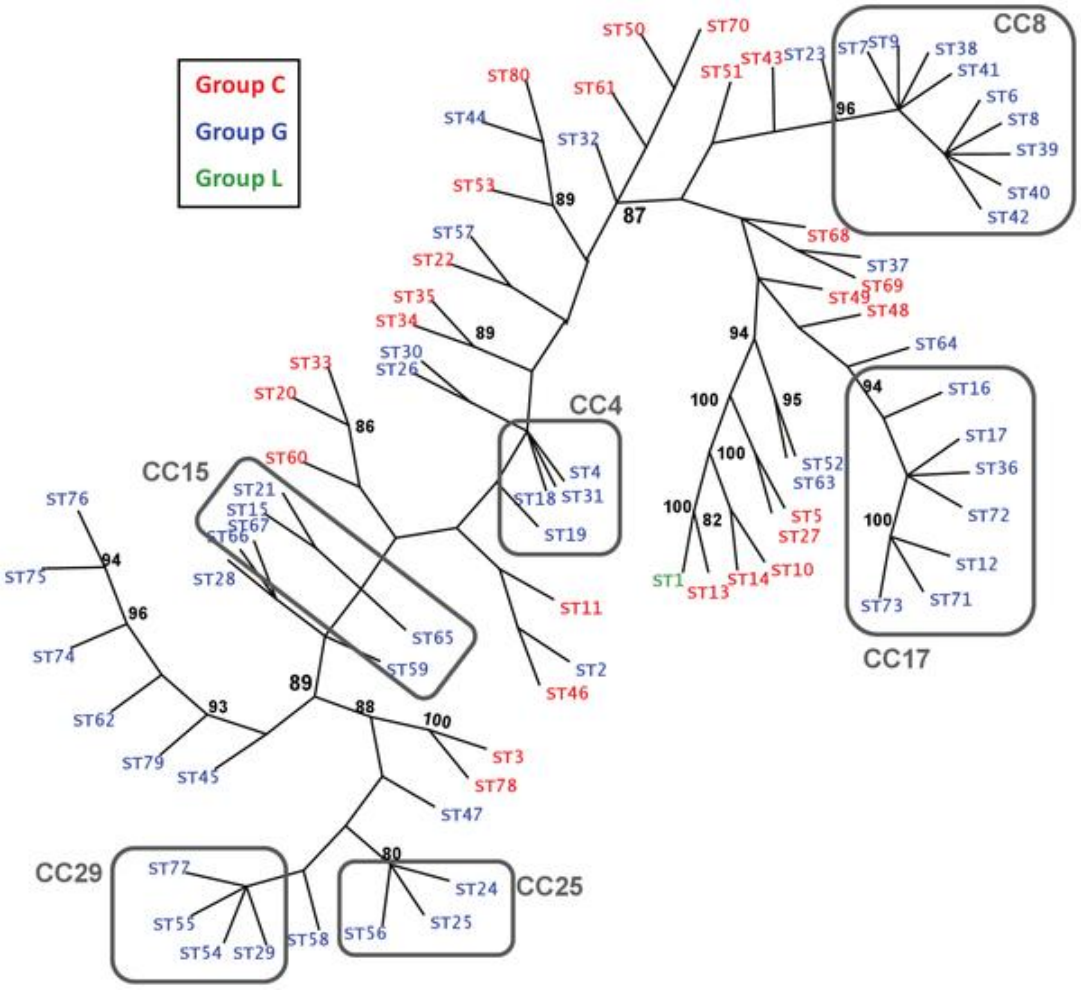

FIGURE 3.3. Maximum parsimony tree of concatenated housekeeping alleles

The housekeeping alleles for each of the 80 STs for SDE were concatenated (3,134 nt positions), and a maximum parsimony tree was constructed. The radial, unrooted phylogenetic tree is shown. Bootstrap values (500 replicates) showing branch support equal or greater than 80% are indicated; bootstrap analysis used a heuristic search and the 50% majority-rule consensus tree is presented. STs representing GCS and GGS are depicted in red and blue, respectively; the single group L isolate (ST1) is depicted in green. CCs having three or more STs are indicated. Characters: 2937 are constant, 56 variable characters are parsimony-uninformative, 141 are parsimony-informative. Consistency index (CI) = 0.3350; CI excluding uninformative characters = 0.2669; retention index (RI) = 0.7926.

contrast, SDE alleles *recP12* and *recP3* seem to result from recombination with GAS alleles ($p = 0.00262$ and $p = 0.00231$, respectively). No significant recombination events between the other GAS and SDE housekeeping alleles (*gtr*, *murI*, *mutS*, *xpt*) were observed using RDP. The SplitsTrees analysis for networked evolution was also used to assess intragenic recombination involving the MLST genes within the SDE population. Statistically significant evidence for recombination (PHI test) was observed for *murI*, *recP*, *xpt* and *atoB* ($p<0.01$).

It is important to note that the recombination detection algorithms do not detect the complete replacement of the analyzed fragment. As pointed out previously, at least one instance of recombination involving the entire fragment of *mutS* analyzed is also suggested by phylogenetic analysis (Figure S3.2). Taken together, these data strongly suggest that intra- and inter-species recombination has occurred for several housekeeping genes.

The largest CC (CC8) contains 11.3% of the total STs, and is within the range of reliable performance of the BURST rules (27). To estimate the relative role of recombination versus mutation in the short-term evolution of SDE, the SLVs linked by goeBURST were examined for genetic differences between the variant alleles of each SLV pair. goeBURST identified 38 primary SLVs. Of these, 31 are predicted to have arisen via recombination, and seven SLVs are predicted to arise through point mutation. Thus, recombination occurred 4.4 times more often than mutation. The per site recombination to mutation (*r/m*) ratio was 20.7. An additional four SLVs, representing alternative ST relationships, are also present in the data set; of these, three SLVs are predicted to arise through recombination. The empirical findings on the relative contribution of recombination versus mutation to the genetic diversification of housekeeping genes are highly consistent with other findings that indicate extensive recombination involving SDE.

DISCUSSION

SDE is increasingly recognized as an important human pathogen that causes disease in many regions of the world. The findings of this report demonstrate that the major genotypes of SDE have an intercontinental distribution. The recovery from all three continents of the likely founder ST of at least five CCs supports a model whereby a few successful clones have undergone extensive migration, followed by genetic diversification. Several of the descendents are also widely disseminated, indicative of subsequent waves of clonal migration.

SDE is largely a commensal species, yet the vast majority of isolates evaluated in this study were recovered from cases of human disease. The mode of person-to-person transmission of SDE has not been well-characterized, and there may be large differences among SDE strains in terms of their virulence properties and ease of transmission to new hosts. It stands to reason that the most widely disseminated clones are probably among the most readily transmitted. Whether transmission is positively linked to virulence is an important question that remains to be established for SDE. Molecular typing of SDE, as provided in this report, provides a framework upon which the question of whether or not subpopulations of SDE strains have heightened virulence can be addressed.

More than half (60%) of the SDE isolates studied have a unique combination of *emm* type, ST and group carbohydrate, indicative of a very high level of genetic diversity within the species. The high level of strain diversity may be a consequence of a high rate of genetic change and/or a very large population size. Nonetheless, molecular typing using only the *emm* gene versus MLST, yields stratifications that are highly discordant, and neither method by itself is satisfactory for defining strains or clones. This finding provides support for a role of extensive horizontal gene transfer and recombination in promoting random associations of *emm* and ST.

Recombination following horizontal gene transfer in SDE is observed at several levels of biological importance, involving core housekeeping genes and the *emm* gene, and perhaps even the genes encoding the group carbohydrate biosynthetic enzymes, albeit at a much lower frequency. Inter-specific gene transfer between SDE and GAS is likely for several of the housekeeping genes. Interestingly, ST3 and ST78 isolates possess the SDE *recP6* and *xpt2* alleles, which are identical to the GAS derived *recP83*

and *xpt4* alleles respectively. Although it is formally possible that the *recP* and xpt alleles were transferred in a single genetic exchange event involving a large genome segment, this hypothesis is unlikely because SDE-like *gki* and *gtr* alleles are positioned in between the *xpt* and *recP loci* on the SDE genome.

Transfer of *emm* genes between SDE and GAS also seem likely to have occurred, but to a much more limited extent, as evidenced only by the prototypical GAS *emm* types *emm57* (25) and *emm12* (28). In addition to *emm* and the housekeeping genes, lateral exchange between GAS and SDE is documented for several other genes, including those encoding a fibronectin-binding protein (29), a DNA gyrase subunit implicated in fluoroquinolone resistance (30), a transcriptional regulator of pilus gene expression (31), and the plasminogen activator streptokinase (32, 33). Thus, SDE and GAS share common gene pools for numerous *loci*.

Intra-species genetic exchange among SDE organisms, involving either the core housekeeping genes or the *emm* gene, has been extensive. Data suggest recombination to be the predominant mechanism of genetic diversification among SDE, occurring four times more often than point mutation in housekeeping *loci*. The per-site recombination to mutation ratio is also greater than 20. More than half of the STs represented by more than one isolate are considered to be *emm* variable STs, and are found in association with at least two distinct *emm* types. ST15 is a particularly successful recipient of *emm* genes originating from multiple SDE donor strains, as evidenced by its recovery in association with seven different *emm* types. Mechanisms that might explain the existence of STs having associations with many distinct *emm* types include possession of genetic machinery that increases their probability for recombinational success, additional accessory genes which facilitate their transmission to new hosts, a high prevalence (eg., ST15 comprises 20 of the 178 isolates) or natural selection favoring the emergence of variants harboring unique *emm* types.

Numerous *emm* gene donor-recipient ST pairs are also represented within the SDE isolate data set, wherein 13 different *emm* types are associated with two or more genetically distant STs; in total, 21 distinct *emm* gene horizontal transfer events are suggested by the data. One *emm* type (*stG643*) is found in association with five distant STs or CCs. The *emm* gene of GAS is part of an ancient pathogenicity island (34). Thus, it will be of interest to determine whether the *stG643* gene is harbored by a functional mobile genetic element.

In GAS, the *emm* gene product (M protein) is a major virulence factor and a primary target of protective immunity by the human host. However, to our knowledge, it remains to be shown that the M protein of SDE prevents phagocytosis in the absence of M type-specific antibody, which is a hallmark feature of GAS. Thus, the relationship between gene replacements involving the *emm* gene, and positive selection arising from host immune pressures, remains speculative for SDE. The M proteins of GAS are also multifunctional proteins, which as a group, exhibits binding for numerous host proteins that include plasminogen, fibrinogen, immunoglobulins and complement regulatory proteins, as recently reviewed by Smeesters *et al.* (35). Like GAS, M proteins of SDE that can bind plasminogen have been identified (36). However, the full extent by which the M proteins of SDE share functional attributes with the M protein of GAS remains to be established. The acquisition of new, M protein-mediated functional activities by a recipient SDE strain could conceivably drive selection for the emergence of novel *emm*-ST combinations among SDE.

Group carbohydrate is synthesized by biosynthetic enzymes whose genes have yet to be characterized. Among SDE, the association between group C versus G carbohydrate and ST approaches 100% concordance. The observed linkage between group carbohydrate and ST is probably a large reflection of short term evolution. An evolutionary history of housekeeping gene exchange between GCS and GGS is evident from the phylogenetic tree of concatenated housekeeping gene sequences. While it is likely that GCS and GGS diverged from a common ancestor, subsequent genetic exchange masks that history, making it difficult to ascertain the extent to which GCS and GGS comprise distinct evolutionary lineages. However, MLST data makes it clear that lateral gene transfer leading to a group carbohydrate switch is a rare event.

The goeBURST population snapshot of SDE (Figure 3.1) differs from that reported for GAS (16). For GAS, only 2.8% of the STs are present in the largest CC, as compared to 11.2% of the SDE STs (i.e., CC8). Based on simulated bacterial populations differing in levels of recombination and diversity generated by mutation (27), the population genetic structure of GAS is best explained by high rates of both recombination and mutation acting on a diverse set of housekeeping genes. For SDE, the recombination rate and housekeeping gene diversity may be similarly high; however the mutation rate may be somewhat lower than in GAS. This proposed genetic structure for the SDE population is supported by correspondence to simulated populations (27),

combined with findings that show a greater than 4-fold excess of recombinational events in the diversification of SLVs.

SDE appears to be among the closest extant relatives of GAS. In general terms, a pathogenic species of bacteria can arise from an organism of lower virulence following acquisition of virulence genes (eg., pathogenicity islands). However, evolution can also flow in the opposite direction, as recently evidenced by the loss of virulence genes and descent of the commensal *Streptococcus mitis* from the pathogen *Streptococcus pneumoniae* (37). Determination of whether the most recent common ancestor of GAS and SDE is more closely related to GAS or to SDE, will probably require more extensive genomic analyses. The molecular typing and characterization of the population biology of SDE in this report provides a foundation for future studies that address the evolution and molecular basis for virulence in SDE.

ACKNOWLEDGMENTS

We thank Awdhesh Kalia for providing genome sequence data that was used in the design of new oligonucleotide primers. We also thank Bernard Beall for providing additional details of previously described isolates (25).

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: D. J. McMillan, D. E. Bessen and M. Ramirez. Performed the experiments: D. J. McMillan, M. Pinho, C. Ford and J. S. Hall. Analyzed the data: D. J. McMillan, D. E. Bessen, M. Pinho, C. Ford, J. S. Hall, J. Melo-Cristino and M. Ramirez. Contributed reagents/materials/analysis tools: D. J. McMillan, D. E. Bessen and M. Ramirez. Wrote the paper: D. J. McMillan, D. E. Bessen, M. Pinho, C. Ford, J. S. Hall, J. Melo-Cristino and M. Ramirez.

REFERENCES

1. **Facklam R**. 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin. Microbiol. Rev. **15**:613–630.

2. **Chhatwal GS, McMillan DJ, Talay SR**. 2006. Pathogenicity factors in group C and G streptococci. In: Fischetti VA, Novick RP, Ferretti JJ, Portnoy DA, Rood JI, editors. Gram-positive pathogens. 2 ed. Washington D. C.: ASM Press. pp. 213–221.

3. **Broyles LN, Van Beneden C, Beall B, Facklam R, Shewmaker PL, Malpiedi P, Daily P, Reingold A, Farley MM. 2009.** Population-based study of invasive disease due to beta-hemolytic streptococci of groups other than A and B. Clin. Infect. Dis. **48**:706–712.

4. **Efstratiou A.** 1997. Pyogenic streptococci of Lancefield groups C and G as pathogens in man. Soc. Appl. Bacteriol. Symp. Ser. **26 Suppl**:72–79.

5. **Pinho MD, Melo-Cristino J, Ramirez M.** 2006. Clonal relationships between invasive and noninvasive Lancefield group C and G streptococci and *emm*-specific differences in invasiveness. J. Clin. Microbiol. **44**: 841–846.

6. **Humar D, Datta V, Bast DJ, Beall B, De Azavedo JCS, Nizet V**. 2002. Streptolysin S and necrotising infections produced by group G streptococcus. Lancet **359**:124–129.

7. **Woo PC, Fung AM, Lau SK, Wong SS, Yuen KY.** 2001. Group G beta-hemolytic streptococcal bacteremia characterized by 16S ribosomal RNA gene sequencing. J. Clin. Microbiol. **39**:3147–3155.

8. **Hashikawa S, Iinuma Y, Furushita M, Ohkura T, Nada T, Torii K, Hasegawa T, Ohta M**. 2004. Characterization of group C and G streptococcal strains that cause streptococcal toxic shock syndrome. J. Clin. Microbiol. **42**:186–192.

9. **Korman TM, Boers A, Gooding TM, Curtis N, Visvanathan K**. 2004. Fatal case of toxic shock-like syndrome due to group C streptococcus associated with superantigen exotoxin. J. Clin. Microbiol. **42**:2866–2869.

10. **Bramhachari PV, Kaul SY, McMillan DJ, Shaila MS, Karmarkar MG, Sriprakash KS**. 2010. Disease burden due to *Streptococcus dysgalactiae* subsp. *equisimilis* (group G and C streptococcus) is higher than that due to *Streptococcus pyogenes* among Mumbai school children. J. Med. Microbiol. **59**:220–223.

11. **McDonald M, Towers RJ, Andrews RM, Carapetis JR, Currie B**. 2007. Epidemiology of *Streptococcus dysgalactiae* subsp. *equisimilis* in tropical communities, Northern Australia. Emerg. Infect. Dis. **13**:1694–1700.

12. **Tanaka D, Isobe J, Watahiki M, Nagai Y, Katsukawa C, Kawahara R, Endoh M, Okuno R, Kumagai N, Matsumoto M, Morikawa Y, Ikebe T, Watanabe H**. 2008. Genetic features of clinical isolates of *Streptococcus dysgalactiae* subsp. *equisimilis* possessing Lancefield's group A antigen. J. Clin. Microbiol. **46:**1526–1529.

13. **Davies MR, McMillan DJ, Beiko RG, Barroso V, Geffers R, Sriprakash KS, Chhatwal GS.** 2007. Virulence profiling of *Streptococcus dysgalactiae* subspecies *equisimilis* isolated from infected humans reveals 2 distinct genetic lineages that do not

segregate with their phenotypes or propensity to cause diseases. Clin. Infect. Dis. **44**:1442–1454.

14. **Ikebe T, Murayama S, Saitoh K, Yamai S, Suzuki R, Isobe J, Tanaka D, Katsukawa C, Tamaru A, Katayama A, Fujinaga Y, Hoashi K, Watanabe H.** 2004. Surveillance of severe invasive group-G streptococcal infections and molecular typing of the isolates in Japan. Epidemiol. Infect. **132**:145–149.

15. **Igwe EI, Shewmaker PL, Facklam RR, Farley MM, van Beneden C, Beall B**. 2003. Identification of superantigen genes *speM*, *ssa*, and *smeZ* in invasive strains of beta-hemolytic group C and G streptococci recovered from humans. FEMS Microbiol. Lett. **229**:259–264.

16. **Bessen DE. 2009.** Population biology of the human restricted pathogen, *Streptococcus pyogenes*. Infect. Genet. Evol. **9:**581–593.

17. **Bessen DE, Lizano S.** 2010. Tissue tropisms in group A streptococcal infection. Future Microbiol. **5:**623-638.

18. **Cleary PP, Kaplan EL, Handley JP, Wlazlo A, Kim MH, Hauser AR, Schlievert PM.** 1992. Clonal basis for resurgence of serious *Streptococcus pyogenes* disease in the 1980s. Lancet **339**:518–521.

19. **Shulman ST, Tanz RR, Dale JB, Beall B, Kabat W, Kabat K, Cederlund E, Patel D, Rippe J, Li Z, Sakota V**. 2009. Seven-year surveillance of North american pediatric group A streptococcal pharyngitis isolates. Clin. Infect. Dis. **49**:78–84.

20. **Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE**. 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between *emm* type and clone. Infect. Immun. **69**:2416–2427.

21. **McGregor KF, Spratt BG, Kalia A, Bennett A, Bilek N, Beall B, Bessen DE.** 2004. Multilocus sequence typing of *Streptococcus pyogenes* representing most known *emm* types and distinctions among subpopulation genetic structures. J. Bacteriol. **186**:4285–4294.

22. **Sakota V, Fry AM, Lietman TM, Facklam RR, Li Z, Beall B.** 2006. Genetically diverse group A streptococci from children in far-western Nepal share high genetic relatedness with isolates from other countries. J. Clin. Microbiol. **44**:2160–2166.

23. **Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan MS, Kunst F, Glaser P, Rusniok C, Crook DW, Harding RM, Bisharat N, Spratt BG**. 2003. Multilocus sequence typing system for group B streptococcus. J. Clin. Microbiol. **41**:2530–2536.

24. **Springman AC, Lacher DW, Wu G, Milton N, Whittam TS, Davies HD, Manning SD**. 2009. Selection, recombination, and virulence gene diversity among group B streptococcal genotypes. J. Bacteriol. **191**:5419–5427.

25. **Ahmad Y, Gertz RE, Li Z, Sakota V, Broyles LN, Van Beneden C, Facklam R, Shewmaker PL, Reingold A, Farley MM, Beall BW**. 2009. Genetic relationships deduced from *emm* and multilocus sequence typing of invasive *Streptococcus dysgalactiae* subsp. *equisimilis* and *S. canis* recovered from isolates collected in the United States. J. Clin. Microbiol. **47**:2046–2054.

26. **Martin DP**. 2009. Recombination detection and analysis using RDP3. Methods. Mol. Biol. **537**:185–205.

27. **Turner KM, Hanage WP, Fraser C, Connor TR, Spratt BG**. 2007. Assessing the reliability of eBURST using simulated populations with known ancestry. BMC Microbiol. **7**: 30.

28. **Simpson WJ, Musser JM, Cleary PP**. 1992. Evidence consistent with horizontal transfer of the gene (*emm12*) encoding serotype M12 protein between group A and group G pathogenic streptococci. Infect. Immun. **60**:1890–1893.

29. **Towers RJ, Gal D, McMillan D, Sriprakash KS, Currie BJ, Walker MJ, Chhatwal GS, Fagan PK**. 2004. Fibronectin-binding protein gene recombination and horizontal transfer between group A and G streptococci. J. Clin. Microbiol. **42**:5357–5361.

30. **Pinho MD, Melo-Cristino J, Ramirez M**. 2010. Fluoroquinolone resistance in *Streptococcus dysgalactiae* subsp. *equisimilis* and evidence for a shared global gene pool with *Streptoccocus pyogenes*. Antimicrob. Agents Chemother. **54**:1769-1777.

31. **Bessen DE, Manoharan A, Luo F, Wertz JE, Robinson DA**. 2005. Evolution of transcription regulatory genes is linked to niche specialization in the bacterial pathogen *Streptococcus pyogenes*. J. Bacteriol. **187**:4163–4172.

32. **Kalia A, Bessen DE**. 2004. Natural selection and evolution of streptococcal virulence genes involved in tissue-specific adaptations. J. Bacteriol. **186**:110–121.

33. **Musser JM, Kapur V, Szeto J, Pan X, Swanson DS, Martin DR**. 1995. Genetic diversity and relationships among *Streptococcus pyogenes* strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease. Infect. Immun. **63**:994–1003.

34. **Panchaud A, Guy L, Collyn F, Haenni M, Nakata M, Podbielski A, Moreillon P, Roten C-AH**. 2009. M-protein and other intrinsic virulence factors of *Streptococcus pyogenes* are encoded on an ancient pathogenicity island. BMC Genomics **10**:198.

35. **Smeesters PR, McMillan DJ, Sriprakash KS**. 2010. The streptococcal M protein: a highly versatile molecule. Trends Microbiol. **8:**275-282.

36. **Ben Nasr A, Wistedt A, Ringdahl U, Sjobring U.** 1994. Streptokinase activates plasminogen bound to human group C and G streptococci through M-like proteins. Eur. J. Biochem. **222**:267–276.

37. **Kilian M, Poulsen K, Blomqvist T, Håvarstein LS, Bek-Thomsen M, Tettelin H, Sørensen UB.** 2008. Evolution of *Streptococcus pneumoniae* and its close commensal relatives. PLoS One **3:**e2683.

38. **Beall B, Facklam R, Thompson T**. 1996. Sequencing *emm*-specific PCR products for routine and accurate typing of group A streptococci. J. Clin. Microbiol. **34**:953–958.

39. **Francisco AP, Bugalho M, Ramirez M, Carrico JA**. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics **10**:152.

40. **Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG**. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186**:1518–1530.

41. **Carrico JA, Silva-Costa C, Melo-Cristino J, Pinto FR, de Lencastre H, Almeida JS, Ramirez M**. 2006. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. J. Clin. Microbiol. **44**:2524–2532.

42. **Librado P, Rozas J**. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**:1451–1452.

43. **Tamura K, Dudley J, Nei M, Kumar S**. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**:1596–1599.

44. **Feil EJ, Enright MC, Spratt BG**. 2000. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. Res. Microbiol. **151**:465–469.

45. **Huson DH, Bryant D**. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23**:254–267.

46. **Bruen TC, Philippe H, Bryant D.** 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics **172**:2665–2681.

SUPPLEMENTAL MATERIAL

TABLE S3.1. Characteristics of SDE isolates in this study

Strain	*emm* type	Group carbohydrate	ST	Clonal Complex[a]	Location	Collection
168554	*stG485*	G	47	37	Portugal	UL
171712	*stG480*	G	38	8	Portugal	UL
220269	*stG2078*	G	15	15	Portugal	UL
223754	*stC839*	C	3	3	Portugal	UL
230631	*stG480*	G	8	8	Portugal	UL
231995	*stC74a*	G	29	29	Portugal	UL
241940	*stC36*	C	50	49	Portugal	UL
273600	*stG166b*	G	65	15	Portugal	UL
299298	*stG643*	G	8	8	Portugal	UL
313247	*stG6*	G	25	25	Portugal	UL
363962	*stG2078*	G	17	17	Portugal	UL
378119	*stC839*	G	15	15	Portugal	UL
380870	*stG480*	G	41	8	Portugal	UL
386041	*stC839*	C	3	3	Portugal	UL
394314	*stG2078*	G	72	17	Portugal	UL
423738	*stG62647*	C	20	20	Portugal	UL
450784	*stG10*	G	15	15	Portugal	UL
460880	*stG10*	G	15	15	Portugal	UL
493188	*stG485*	C	69	s	Portugal	UL
542567	*stG6*	G	62	s	Portugal	UL
618280	*EMM57*	G	57	s	Portugal	UL
SH0004	*stG6792*	G	4	31	Portugal	UL
SH0015	*stG6*	G	25	25	Portugal	UL
SH0032	*stG166b*	G	15	15	Portugal	UL
SH0102	*stG2078*	G	17	17	Portugal	UL
SH0107	*stG643*	G	52	52	Portugal	UL
SH0110	*stG6*	G	25	25	Portugal	UL
SH0113	*stG6792*	G	4	31	Portugal	UL
SH0124	*stG6792*	G	4	31	Portugal	UL
SH0218	*stG245*	G	15	15	Portugal	UL
SH0254	*stG485*	C	69	s	Portugal	UL
SH0257	*stC6979*	C	80	s	Portugal	UL
SH0259	*stG652*	G	71	17	Portugal	UL
SH0275	*stG485*	G	55	29	Portugal	UL
SH0330	*stC36*	C	49	49	Portugal	UL
SH0336	*stG5420*	G	25	25	Portugal	UL
G121	*stC74a*	G	29	29	Australia	QIMR
G122	*stC74a*	G	29	29	Australia	QIMR
GCS10128	*stC1400*	C	46	s	Australia	QIMR
GCS2816	*stG62647*	C	20	20	Australia	QIMR
GCS6894	*stG62647*	C	20	20	Australia	QIMR
GCS6929	*stG62647*	C	20	20	Australia	QIMR
GGS075	*stG166b*	G	56	25	Australia	QIMR
GGS101	*stG643*	G	12	17	Australia	QIMR
GGS10b	*stG6*	G	44	44	Australia	QIMR
GGS11172	*stC74a*	G	29	29	Australia	QIMR
GGS11543	*stG643*	G	12	17	Australia	QIMR
GGS120	*stG4831*	G	74	74	Australia	QIMR
GGS19	*stC1400*	G	64	s	Australia	QIMR
GGS2	*stG10*	G	15	15	Australia	QIMR
GGS24	*stG6*	G	44	44	Australia	QIMR
GGS430	*stG643*	G	12	17	Australia	QIMR
GGS463	*stG10*	G	15	15	Australia	QIMR
GGS539813	*stC74a*	G	29	29	Australia	QIMR
GGS540048	*stG485*	G	29	29	Australia	QIMR
GGS545448	*stG10*	G	15	15	Australia	QIMR
GGS569	*stC6979*	G	54	29	Australia	QIMR
GGS592	*stG480*	G	8	8	Australia	QIMR

TABLE S3.1. (*cont*)

Strain	*emm* type	Group carbohydrate	ST	Clonal Complex	Location	Collection
GGS9225	*stC74a*	G	29	29	Australia	QIMR
GGS985	*stC1400*	G	66	15	Australia	QIMR
GGSRHD	*stG480*	G	39	8	Australia	QIMR
MD01	*stG11*	G	8	8	Australia	QIMR
MD02	*stG10*	G	15	15	Australia	QIMR
MD03	*stG2078*	G	17	17	Australia	QIMR
MD031	*stC74a*	G	17	17	Australia	QIMR
MD04	*stG6*	G	25	25	Australia	QIMR
MD05	*stG6*	G	63	52	Australia	QIMR
MD06	*stC74a*	G	29	29	Australia	QIMR
MD07	*stG6*	G	58	s	Australia	QIMR
MD08	*stG2078*	G	17	17	Australia	QIMR
MD09	*stG10*	G	15	15	Australia	QIMR
MD122	*stC74a*	G	29	29	Australia	QIMR
MD136	*stG10*	G	15	15	Australia	QIMR
MD163	*stG652*	G	63	52	Australia	QIMR
MD227	*stC6979*	C	20	20	Australia	QIMR
MD248	*stC6979*	G	54	29	Australia	QIMR
MD296	*stG10*	G	15	15	Australia	QIMR
MD411	*stC1400*	G	66	15	Australia	QIMR
MD499	*stC6979*	G	54	29	Australia	QIMR
MD504	*stC839*	C	3	3	Australia	QIMR
MD605	*stG62647*	C	20	20	Australia	QIMR
MD633	*stG643*	G	73	17	Australia	QIMR
MD722	*stC74a*	G	29	29	Australia	QIMR
MD805	*stG652*	G	59	15	Australia	QIMR
MD834	*stC36*	G	4	31	Australia	QIMR
MD872	*stG62647*	C	60	s	Australia	QIMR
MD921	*stC6979*	G	54	29	Australia	QIMR
MD934	*stG5420*	G	25	25	Australia	QIMR
NS1121	*stG4831*	G	75	74	Australia	QIMR
NS542	*stG652*	G	15	15	Australia	QIMR
NS752	*stG6*	G	44	44	Australia	QIMR
GCS01ny	*stC1400*	C	3	3	Other	NYMC
GCS04ny	*stC5344*	C	43	s	Other	NYMC
GCS07ny	*stG6792*	C	51	s	Other	NYMC
GCS08ny	*stC5345*	C	53	44	Other	NYMC
GCS10ny	*stC74a*	C	70	s	Other	NYMC
GCS11ny	*stC839*	C	78	3	Other	NYMC
GGS12ny	*stG11*	G	8	8	Other	NYMC
GGS13ny	*stG480*	G	8	8	Other	NYMC
GGS15ny	*stC74a*	G	29	29	Other	NYMC
GGS16ny	*stG485*	G	29	29	Other	NYMC
GGS19ny	*stG11*	G	42	8	Other	NYMC
GGS24ny	*stG4974*	G	79	s	Other	NYMC
GGS22ny	*stG4831*	G	76	s	Other	NYMC
GGS 124	*stG480*	G	67	15	Other	other
DSM6176	*stG652*	C	61	s	Other	other
GCS02ny	*stC839*	C	3	3	USA	NYMC
GCS03ny	*stG653*	C	3	3	USA	NYMC
GCS05ny	*stG643*	C	48	s	USA	NYMC
GCS06ny	*stC36*	C	49	49	USA	NYMC
GCS09ny	*stC36*	C	68	s	USA	NYMC
GGS14ny	*stG10*	G	15	15	USA	NYMC
GGS17ny	*stG480*	G	38	8	USA	NYMC
GGS18ny	*stG480*	G	40	8	USA	NYMC
GGS20ny	*stC36*	G	45	s	USA	NYMC
GGS21ny	*stG6*	G	52	52	USA	NYMC
GGS23ny	*stC74a*	G	77	29	USA	NYMC

[a] 's' denotes singleton

TABLE S3.2. PCR primer pairs used for MLST in the study

Locus	**Forward primer**	**Forward primer sequence (5' to 3')**	**Reverse primer**	**Reverse primer sequence (5' to 3')**
gki	Gkigc-up	GGAATTGGTATGGGATCACCAGGAGC	Gkigc-dn	AATTCTCCTGCTGCTGACAC
gtr	Gtrgc-up	GCACAAGTATTATGGGCACA	Gtrgc-dn	CACGGTCTGCGACTTC
gtr	Gtrgc-up2	AAATATTATGGGCAAAACGAGGTG	Gtrgc-dn2	CTTCCACAATAACGCCGCCATCCATA
murI	Murigc-up	GACCTGCTGAGCAAATTAGAGAATACACATGGG	Murigc-dn	CAGGACTTGCCGTTGTGTAAAAATGGTG
murI	Murigc-up2	TGCTGACTCAAAATGTTAAAATGATTG	Murigc-dn2	GATGATAATTCACCGTTAATGTCAAAATAG
mutS	MutSgc-up	GAAGAGTCATCTAGTTTAGAATACGAT	MutSgc-dn	AGAGAGTTGTCACTTGCGCGTTTGATTGCT
recP	RecPgc-up	GCAAATTCTGGACACCCAGG	RecPgc-dn	CTTTCACAAGGATATGTTGCC
xpt	Xptgc-up	TTACTTGAAGAACGCATCTTA	Xptgc-dn	ATGAGGTCACTTCAATGCCC
atoB	AtoBgc-up	ACGTTGCTCAGAAATATGGCAT	AtoBgc-dn	AAAGTGTTGCTAGTCCTCTGGTTAC

TABLE S3.2. ***(cont)***

Locus	**Nucleotide location in GGS_124[a]**	**Primers used by[b]**
gki	1473922-1473425	UL, NYMC, QIMR
gtr	1450465-1450016	UL, QIMR
gtr	1450465-1450016	NYMC
murI	389637-390074	UL, NYMC
murI	389637-390074	QIMR
mutS	2031544-2031140	UL, NYMC, QIMR
recP	1668713-1668255	UL, NYMC, QIMR
xpt	882848-883297	UL, NYMC, QIMR
atoB	1634373-1634806	UL, NYMC, QIMR

[a] Based on nucleotide position in the GGS_124 genome (GenBank number AP010935)

[b] UL, University of Lisbon; NYMC, New York Medical College; QIMR, Queensland Institute of Medical Research.

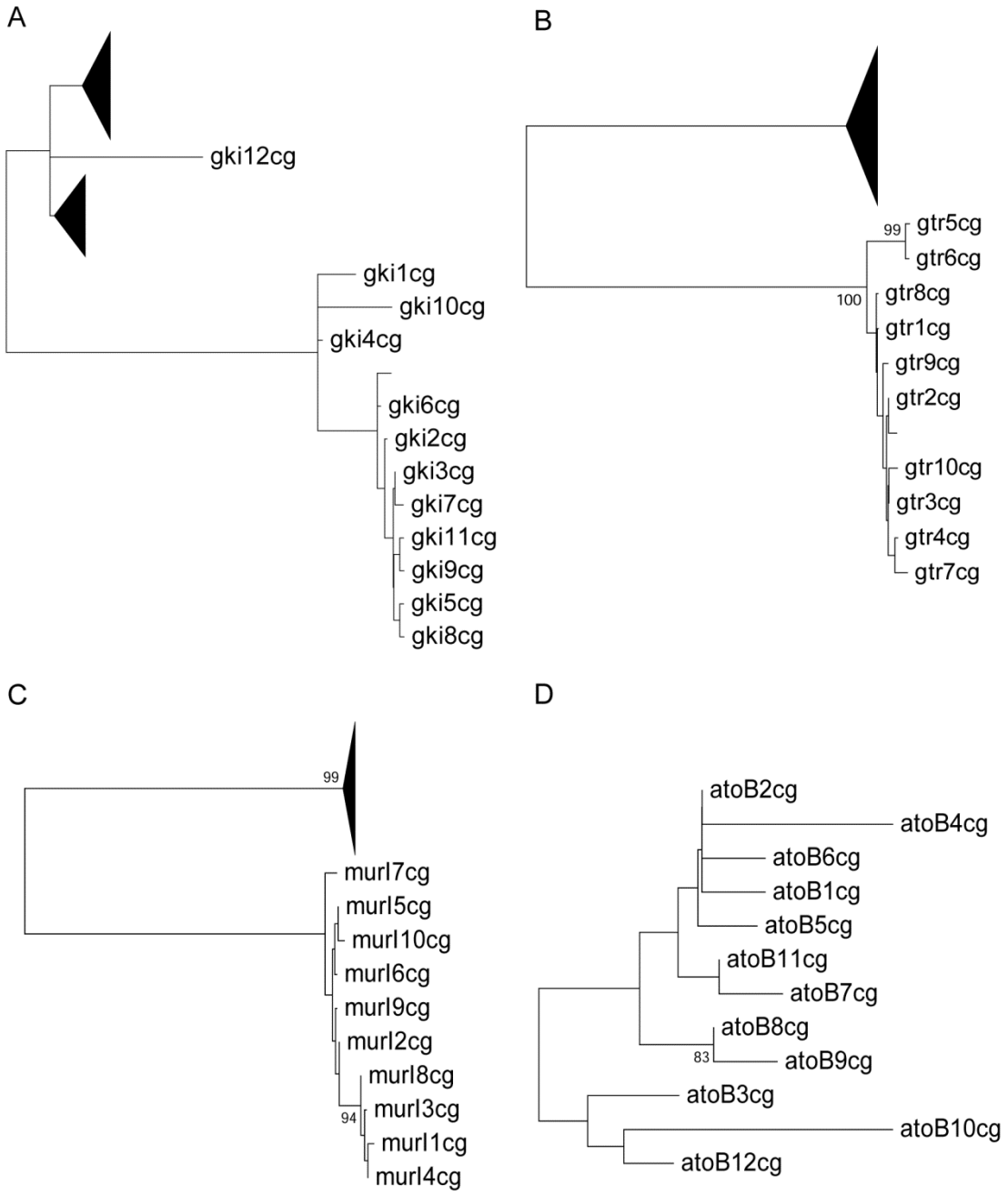

FIGURE S3.1. Evolutionary history of *gki* (A), *gtr* (B), *murI* (C) and *atoB* (D) alleles from SDE and GAS.

Evolutionary relationships were inferred using the Neighbour-Joining (NJ) method, and evolutionary distances calculated using the Jukes-Cantor method. Branches with bootstrap support (n = 1000) greater than 80% are shown next to their respective branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Only names of the GGS alleles are shown. Phylogenetic analyses were conducted in MEGA4.

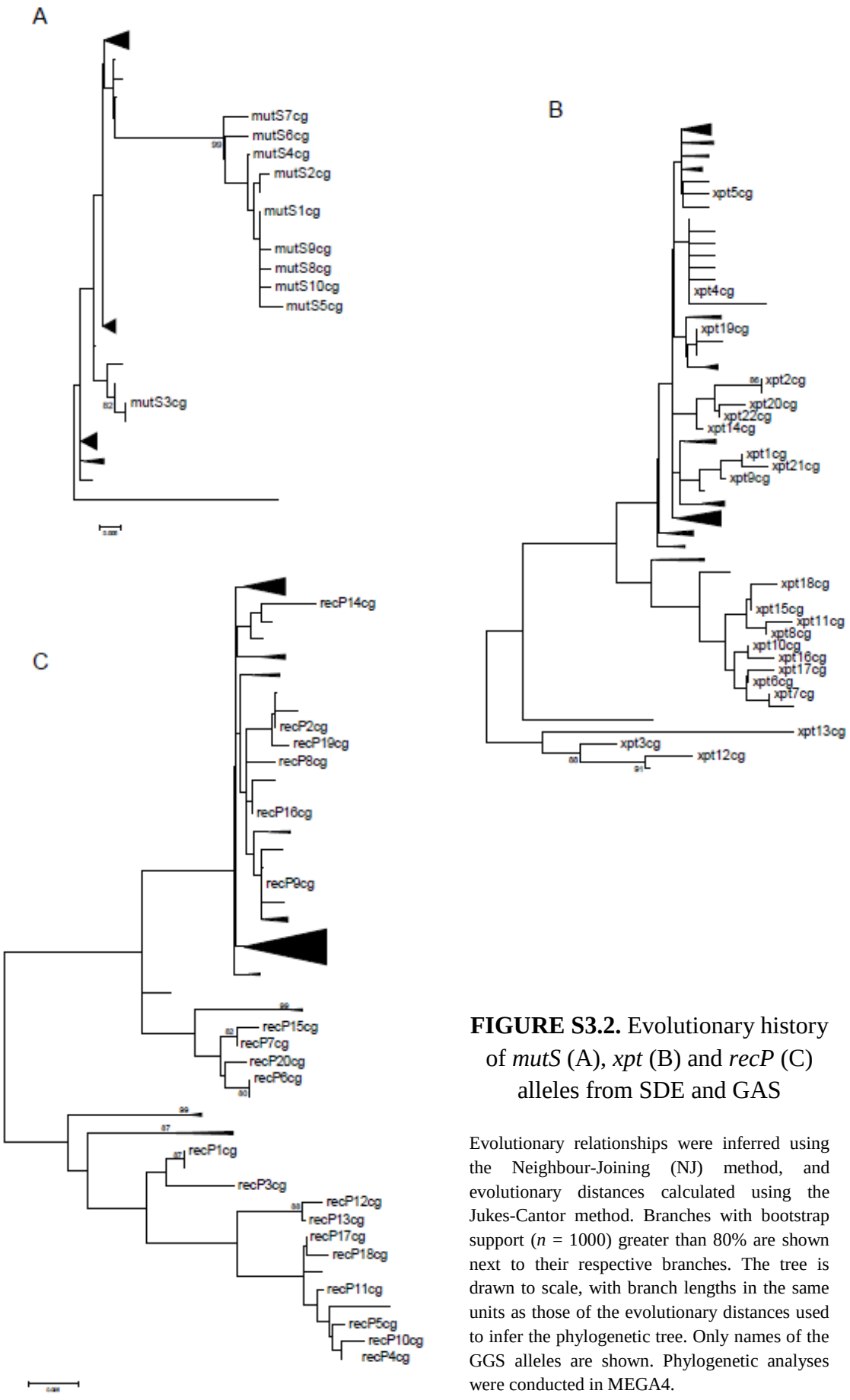

FIGURE S3.2. Evolutionary history of *mutS* (A), *xpt* (B) and *recP* (C) alleles from SDE and GAS

Evolutionary relationships were inferred using the Neighbour-Joining (NJ) method, and evolutionary distances calculated using the Jukes-Cantor method. Branches with bootstrap support (n = 1000) greater than 80% are shown next to their respective branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Only names of the GGS alleles are shown. Phylogenetic analyses were conducted in MEGA4.

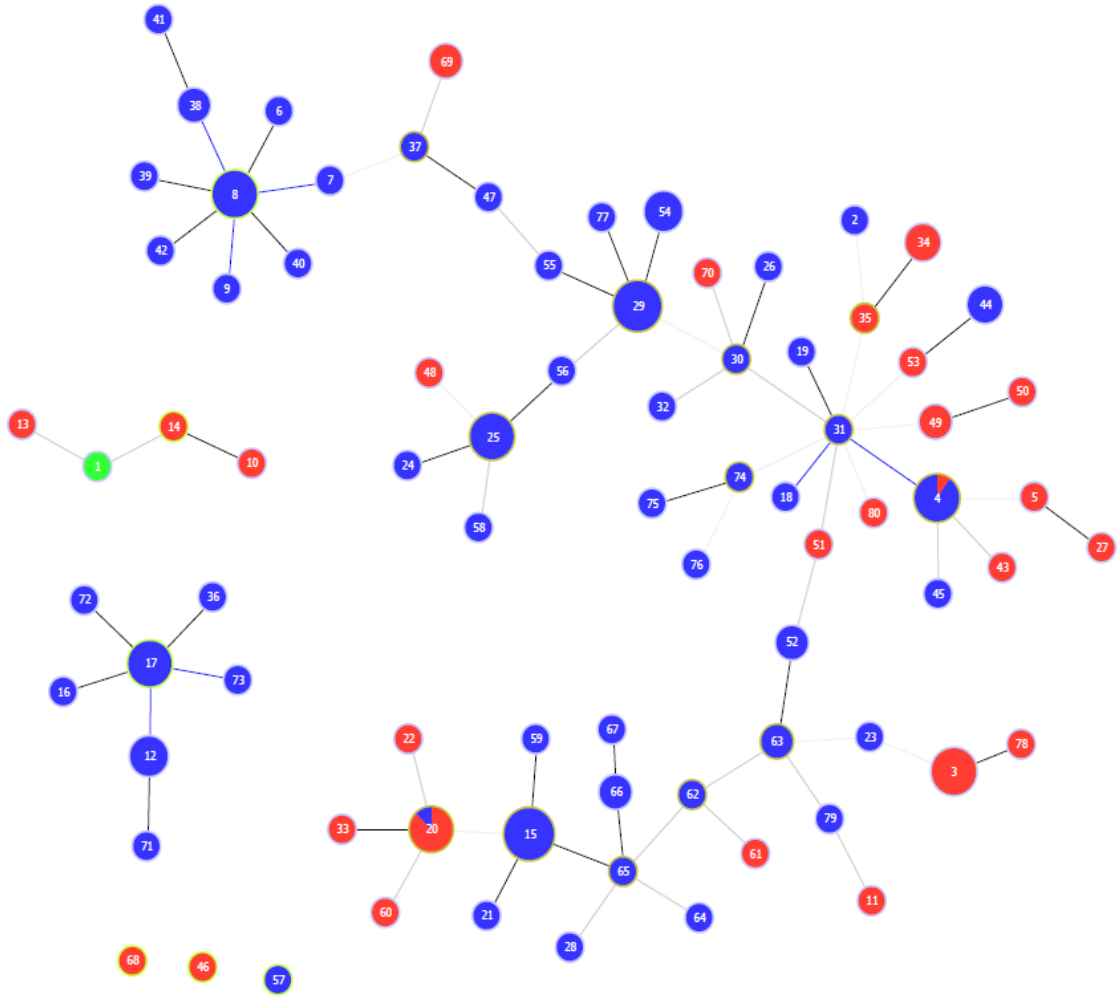

FIGURE S3.3. goeBURST diagram of the relationships between 178 global SDE isolates grouped up to TLV

The size of each circle is proportional to the number of isolates with that particular ST in a logarithmic scale. STs differing up to three alleles (triple-locus-variants - TLVs) are linked by straight lines. Black lines link STs differing at a single gene. Intermediate grey lines link STs differing at a two genes. Light grey lines link STs differing at a three genes. Blue circles represent isolates that have the group G carbohydrate. Red circles represent isolates expressing the group C carbohydrate. Whenever isolates of the same ST have different group carbohydrates, the number of isolates bearing the same carbohydrate is proportional to the respective color. The green circle represents the single isolate expressing the group L carbohydrate. The proposed founders of particular clusters are indicated by a light green outer circle. The sub-founders (defined has having links to three or more STs) are indicated by dark green outer circles.

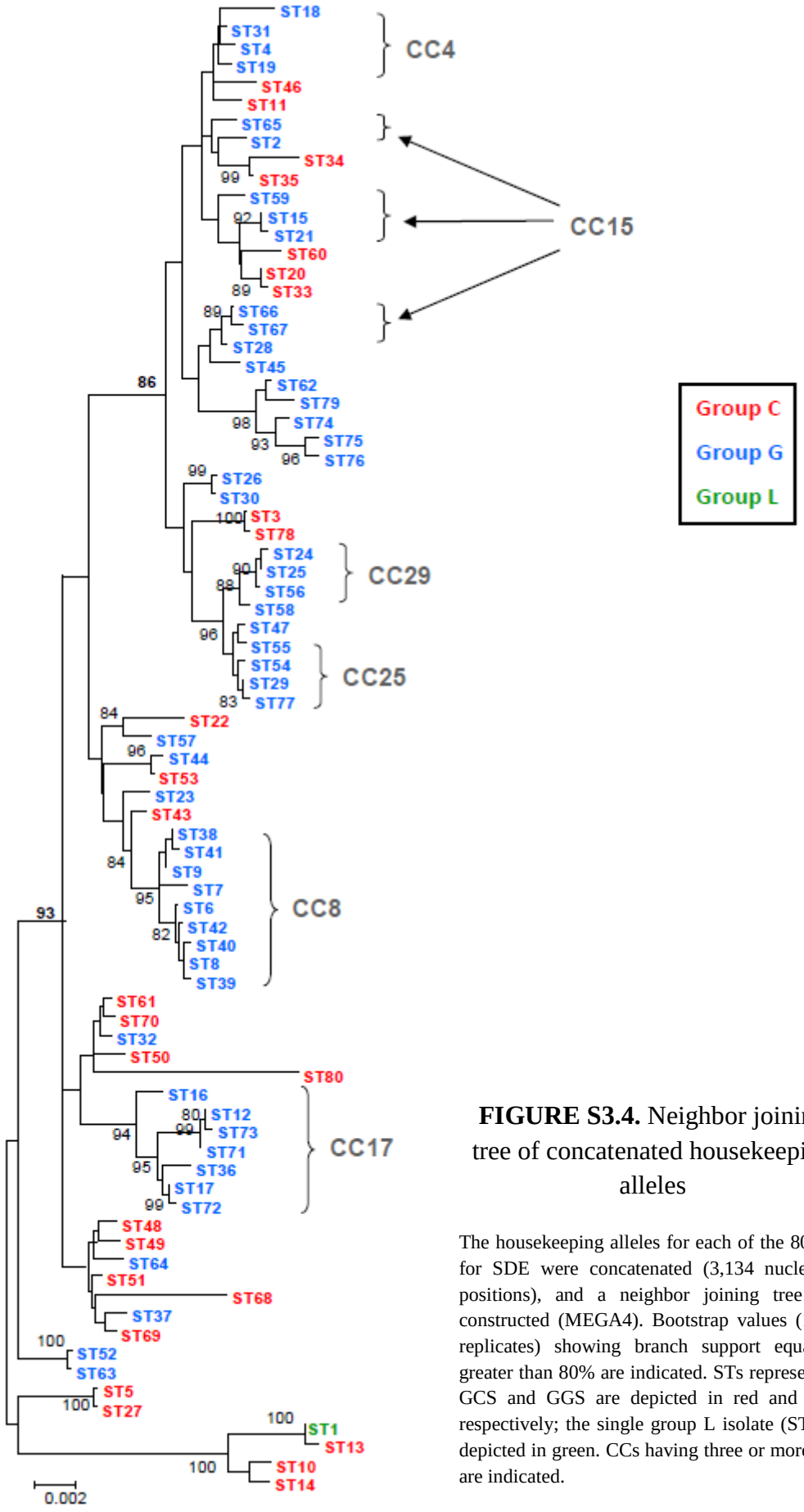

FIGURE S3.4. Neighbor joining tree of concatenated housekeeping alleles

The housekeeping alleles for each of the 80 STs for SDE were concatenated (3,134 nucleotide positions), and a neighbor joining tree was constructed (MEGA4). Bootstrap values (1,000 replicates) showing branch support equal or greater than 80% are indicated. STs representing GCS and GGS are depicted in red and blue, respectively; the single group L isolate (ST1) is depicted in green. CCs having three or more STs are indicated.

CHAPTER 4

Fluoroquinolone resistance in *Streptococcus dysgalactiae* subsp. *equisimilis* and evidence for a shared global gene pool with *Streptococcus pyogenes*

The contents of this chapter were presented in:

Pinho MD, Melo-Cristino J, Ramirez M, and the Portuguese Group for the Study of Streptococcal Infections

2010. Antimicrobial Agents and Chemotherapy **54**:1769-1777

INTRODUCTORY NOTE

The unusual high resistance rate to levofloxacin (12.1%) in SDE that prompted the study described in this chapter was found during a survey of the antimicrobial susceptibility of GCGS isolates recovered from human infections in Portugal. The initial aims of this broader study were to determine the antimicrobial susceptibility patterns and to define whether specific resistance traits were associated with distinct clonal lineages present in the population. The main findings of this study are given below and contextualize the work on levofloxacin resistance described in the main body of this chapter and that resulted in the publication described.

Antimicrobial susceptibility of SDE isolates from human infections in Portugal

A collection of 314 SDE isolates recovered from invasive and non-invasive human infections were collected from June 1998 to December 2005 from 13 laboratories in Portugal (detailed information on the isolates is given in the Materials and Methods section of this chapter).

Susceptibility testing to penicillin, cefotaxime, erythromycin, clindamycin, tetracycline, chloramphenicol, levofloxacin, vancomycin, streptomycin, gentamicin, quinupristin-dalfopristin and linezolide was performed by disk diffusion according to the CLSI guidelines (11). The macrolide, lincosamide and streptogramin B (MLS_B) resistance phenotypes $cMLS_B$ (constitutive), $iMLS_B$ (inducible) and M, were also detected as described in CLSI guidelines (11). The minimum inhibitory concentration (MIC) of penicillin was determined for all isolates, while MIC testing of erythromycin, clindamycin, levofloxacin, streptomycin and gentamicin was performed on resistant isolates to these antibiotics, using E-test strips (bioMérieux, Marcy-l'Etoile, France) and the CLSI interpretative criteria (11). The breakpoints used for both disk diffusion testing and MIC determination for gentamicin and streptomycin were those recommended for detection of high-level resistance in *Enterococcus* spp., given that CLSI breakpoints for aminoglycosides are not available for streptococci.

All 314 SDE isolates analyzed in this study were susceptible to penicillin. The MIC range was 0.006 µg/mL to 0.047 µg/mL, with MIC_{50} = 0.012 µg/mL and MIC_{90} = 0.016 µg/mL. All isolates were also susceptible to cefotaxime, vancomycin, quinupristin-dalfopristin and linezolide by the disk diffusion test.

Overall, 175 (55.7 %) isolates were resistant to at least one of the antibiotics tested (153 Lancefield group G and 22 group C). Tetracycline resistance was expressed by 127 isolates (40.4%), followed by resistance to MLS_B antibiotics (22.0%, *n* = 69), fluoroquinolones (12.1%, *n* = 38) and aminoglycosides (3.5%, *n* = 11). A single isolate was found to be resistant to chloramphenicol. Table 4.1 summarizes the resistance patterns observed.

TABLE 4.1 Resistance patterns of 314 SDE isolates recovered from human infections in Portugal (1998 to 2005)

Antimicrobial agent	Resistance patterns [no. of isolates (%)][a]			MIC range of resistant strains (µg/mL)[b]
	Resistant	Intermediate	Susceptible	
Tetracycline	127 (40.4)	29 (9.2)	158 (50.3)	nd
Erythromycin	69 (22.0)	0 (0)	245 (78.0)	4 to >256 µg/mL
Clindamycin	37 (11.8)	0 (0)	277 (88.2)	1 to >256 µg/mL
Levofloxacin	38 (12.1)	4 (1.3)	272 (86.6)	6 to >32 µg/mL
Streptomycin	11 (3.5)	na	303 (96.5)	>1024 µg/mL
Gentamicin	2 (0.6)	na	312 (99.4)	>1024 µg/mL
Chloramphenicol	1	0 (0)	313	nd

[a] na, not applicable. For the antimicrobial classes for which MIC testing was performed the MIC results were considered definitive. For aminoglycoside high-level resistance, only the resistant and susceptible categories are considered.

[b] nd, not determined.

A total of 127 SDE isolates were resistant to tetracycline, while 29 displayed intermediate resistance as determined by the disk diffusion test. Thus, tetracycline resistance was the most expressive in our study as it was found in over 40% of the isolates.

Resistance to erythromycin was above 20% in our SDE collection. Among the 69 isolates that were resistant to erythromycin, the MLS_B-phenotype of resistance was the most commonly found, being expressed by 64 (92.8%) isolates (MIC range, 4 to >256 µg/mL). Among these isolates, 37 displayed constitutive clindamycin resistance ($cMLS_B$). All $cMLS_B$ isolates had a clindamycin MIC >256 µg/mL, except one isolate for which the MIC to clindamycin was 1 µg/mL (erythromycin MIC = 8 µg/mL). 27 isolates showed inducible resistance to clindamycin (MIC range, 0,125 to 0,5 µg/mL). Most of MLS_B resistant isolates had a MIC to erythromycin >256 µg/mL, but 22 isolates, comprising both constitutive and inducible phenotypes, presented MICs that

ranged from 48 µg/mL to as low as 4µg/mL. The M-phenotype was found to be expressed by only 5 isolates (MIC range, 4 to 8 µg/mL).

Twelve isolates had a resistant or inconclusive disk diffusion test result for the aminoglycosides as defined in CLSI guidelines for enterococci (11) and resistance was confirmed for all but one isolate by MIC testing. Only two isolates had a MIC of gentamicin >1024 µg/mL, hence being classified as high-level resistant to this antibiotic (11). These two resistant isolates plus nine isolates that were susceptible to gentamicin (MIC range, 6 to 64 µg/mL), all showed MICs to streptomycin >1024 µg/mL.

The results of the molecular characterization of all 314 SDE isolates carried out by using PFGE and *emm* typing are detailed in the main body of the chapter (see Table 4.2). Briefly, resistant isolates to at least one antimicrobial class were present in each PFGE cluster and with the exception of *stG62647* (n = 11), all *emm* types found in more than two isolates included resistant isolates. The two biggest PFGE clusters, H_{88} and J_{77}, included mostly susceptible isolates and were not associated with resistance to any specific antimicrobial class. On the other hand, tetracycline resistant isolates in PFGE cluster C_{56}, mostly of *emm* types *stG10* and *stG166b* (the last one also associated with levofloxacin resistance), and macrolide resistant isolates in cluster B_{11} (mostly *stG2078*) were the most notoriously resistant clonal lineages. In general, resistant isolates were not separated from susceptible isolates presenting the same *emm* type by PFGE. Common to most of the prevalent *emm* types in our study was their association mostly into one PFGE cluster, where the majority of susceptible isolates grouped. Thus, apart from the exceptions described above, the observed resistance to the several antimicrobial agents tested, had a polyclonal origin and could not be explained by the occurrence of resistant clonal lineages with increased prevalence.

The finding in our work of a high rate of resistance to levofloxacin (12.1%) was surprising since it has no parallel in the available GCGS literature. These observations prompted the clarification of the resistance mechanisms involved and the detailed study of the SDE population dynamics in Portugal which could explain such a high rate of resistance.

ABSTRACT

Quinolone resistance is an emerging problem in *Streptococcus pyogenes* (Lancefield group A streptococcus [GAS]), and recombination with *Streptococcus dysgalactiae* DNA has been implicated as a frequent mechanism leading to resistance. We have characterized a collection of *S. dysgalactiae* subsp. *equisimilis* (SDE) isolates responsible for infections in humans (*n* = 314) and found a high proportion of levofloxacin-resistant isolates (12%). Resistance was associated with multiple *emm* types and genetic lineages, as determined by pulsed-field gel electrophoretic profiling. Since we could not find evidence for a role of efflux pumps in resistance, we sequenced the quinolone resistance-determining regions of the *gyrA* and *parC* genes of representative resistant and susceptible isolates. We found much greater diversity among the *parC* genes (19 alleles) than among the *gyrA* genes (5 alleles). While single mutations in either GyrA or ParC were sufficient to raise the MIC so that the strains were classified as intermediately resistant, higher-level resistance was associated with mutations in both GyrA and ParC. Evidence for recombination with GAS DNA was found in some *parC* alleles, but this was not exclusively associated with resistance. Our data support the existence of a common reservoir of genes conferring quinolone resistance shared between SDE and GAS, while no recombination with the animal pathogen *S. dysgalactiae* subsp. *dysgalactiae* could be found.

INTRODUCTION

Streptococcus dysgalactiae subsp. *equisimilis* (SDE) is a beta-hemolytic, large-colony-forming species (diameter, >0.5 mm) presenting either the Lancefield group C or the Lancefield group G antigen (15). Among the members of this group, this species is the one the most commonly reported to be a cause of human infections worldwide (27, 33, 41, 45) and was the only one previously detected among Lancefield group C streptococci (GCS) and Lancefield group G streptococci (GGS) causing infections in humans in Portugal (37). Although SDE may be part of the normal human microbiota (8), it has also been identified to be the causative agent of infections of the respiratory tract, infections of the skin and soft tissue, as well as life-threatening infections, such as endocarditis and bacteremia (37, 41, 45). More recently, an increasing number of reports have described the association of these organisms with streptococcal syndromes typically caused by *Streptococcus pyogenes* (a Lancefield group A streptococcus [GAS]), such as streptococcal toxic shock syndrome (22) and acute rheumatic fever (21). Several studies point to a close relationship between SDE and GAS, with multiple common virulence (40) and antibiotic resistance (17) determinants being found in both species.

Fluoroquinolones (FQs) are nowadays valid therapeutic options for the treatment of streptococcal infections (24). However, the emergence of resistance to this antimicrobial class is a global concern and resistance is increasingly being reported among several streptococcal species (13, 26, 30), including SDE (5). FQs act by inhibiting the bacterial DNA gyrase and DNA topoisomerase IV (24). Alteration of the target enzymes is the more widespread mechanism of resistance in streptococci and the only one that confers high-level resistance to FQs (24), but active efflux was also shown to confer low-level resistance in *Streptococcus pneumoniae* (4, 9) and in other viridans group streptococci (20), but this mechanism was not detected in GAS (30).

Both DNA gyrase and topoisomerase IV are tetramers of two different subunits, subunits GyrA and GyrB and subunits ParC and ParE, respectively (24). Previous studies have shown that FQ resistance arises from amino acid changes in the structurally similar proteins GyrA and ParC, while alterations in GyrB and ParE seem to be less relevant (4, 24). Mutations in the DNA sequences of *gyrA* and *parC* that confer resistance to FQs are found in the so-called quinolone resistance-determining regions

(QRDRs) of these genes (34). Mutations in the *gyrA* and *parC* QRDRs are thought to occur spontaneously in bacterial populations, and this is considered the main mechanism of the emergence of FQ resistance (2, 24).

Evidence for the horizontal transfer of the QRDR between streptococci has been found, and this mechanism may also play a role in the rise in the incidence of FQ resistance (3, 14, 16, 38). One study suggested that *S. dysgalactiae* could act as the donor species in the emergence of FQ-resistant GAS (38), and more recently, a considerable proportion of the GAS isolates from Belgium presented characteristics suggestive of the acquisition of their QRDR of *parC* from SDE (14). However, such studies have been hampered by the small number of *S. dysgalactiae parC* and *gyrA* sequences available, which limits the ability to assess the contribution of FQ resistance in this species to a putative QRDR gene pool common among beta-hemolytic streptococci.

We characterized the molecular mechanisms of levofloxacin resistance in isolates of SDE responsible for infections in humans in Portugal. The clonal structure of the isolates was evaluated by using two well-established typing schemes: *emm* typing and SmaI macrorestriction profiling by pulsed-field gel electrophoresis (PFGE) (37). This approach allowed the identification of putative susceptible ancestors of resistant isolates and evaluation of the diversity of the genetic backgrounds of FQ-resistant isolates. We found that FQ resistance emerged by alteration of the target genes in diverse genetic lineages of SDE, as previously shown for other streptococci (2), suggesting that mutation and selection are the primary drivers of resistance. The analysis of the QRDRs of a large number of genetically unrelated isolates allowed us to identify evidence of the frequent ongoing horizontal transfer of the QRDR of *parC* between SDE and GAS, indicating the existence of a gene pool common between these streptococcal species.

MATERIALS AND METHODS

Bacterial isolates

A collection of 314 beta-hemolytic, large-colony-forming (diameter, >0.5 mm) isolates of GGS (*n* = 249) and GCS (*n* = 65) responsible for infections in humans were collected from June 1998 to December 2005. The distribution of isolates over the study period was as follows: 11 in 1998, 28 in 1999, 20 in 2000, 28 in 2001, 29 in 2002, 70 in 2003, 50 in 2004, and 78 in 2005. The sources of the isolates were the following: blood (*n* = 34), other normally sterile sites (*n* = 11), pharyngeal exudate (*n* = 67), sputum (*n* = 37), skin and soft tissue (*n* = 155), vaginal exudate (*n* = 5), and urine (*n* = 5). The strains were recovered in 13 laboratories, located throughout Portugal, that were asked to submit all nonduplicate GCS and GGS isolates associated with human infections. Partial descriptions of some of these isolates were published previously (37).

Species identification

Strains were identified to the genus level by the submitting laboratory. Upon receipt, the Lancefield group was confirmed by a commercial latex agglutination technique (Slidex Strepto kit; bioMérieux, Marcy l'Etoile, France). All further studies were carried out at the Instituto de Microbiologia, Faculdade de Medicina, Universidade de Lisboa. Beta-hemolysis and colony size were confirmed after overnight incubation at 37°C in tryptic soy agar (Oxoid, Hampshire, England) supplemented with 5% (vol/vol) defibrinated sheep blood. Further identification to the species level was done for all non-levofloxacin-susceptible isolates (levofloxacin MICs > 2 μg/mL [11]) and selected susceptible isolates used for comparative analysis by using the API 20 Strep system (bioMérieux). We had previously confirmed the efficacy of this system for the identification of SDE strains by using 16S rRNA gene sequencing as the "gold standard" (37).

Molecular typing

Isolates were characterized by *emm* typing and SmaI PFGE macrorestriction profiling, as described previously (37). The PFGE patterns were compared by using Bionumerics software (Applied Maths, Sint-Martens-Latem, Belgium) to create dendrograms by the unweighted-pair group method with arithmetic averages

(UPGMA). The Dice similarity coefficient was used with optimization and position tolerance settings of 1.0 and 1.5, respectively. Clusters were defined as groups of isolates ($n \geq 2$) presenting profiles with ≥80% similarity in the dendrogram (37).

Simpson's index of diversity (SID) and 95% confidence intervals (CIs) were calculated as described elsewhere (10).

Antimicrobial susceptibility testing and efflux screening

Testing for susceptibility to levofloxacin was done for all 314 isolates by disk diffusion, according to the guidelines of the CLSI (11). Testing of the levofloxacin MIC for nonsusceptible isolates and selected susceptible isolates was performed with Etest strips (AB Biodisk, Solna, Sweden) and by use of the interpretative criteria of the CLSI (11). Nonsusceptible isolates were also tested for active efflux by the agar dilution method on Mueller-Hinton agar (Oxoid, Hampshire, England) supplemented with 5% sheep blood in the presence of levofloxacin or ciprofloxacin with or without 10 μg/mL of reserpine (Sigma-Aldrich, Steinheim, Germany), as described previously (9).

Analysis of DNA sequence conferring levofloxacin resistance

The segments of the *gyrA* and *parC* genes containing the QRDRs were amplified and sequenced by using the primers previously proposed for use for this purpose (26). The nucleotide sequences and the deduced amino acid sequences of the *gyrA* and *parC* QRDRs were determined for 56 SDE isolates which exhibited different levels of resistance to levofloxacin and which represented the main PFGE clusters and *emm* types identified in this work.

Phylogenetic analysis of *gyrA* and *parC*

Sequence alignments were performed manually. MEGA (version 4) software (42) was used to construct phylogenetic trees by using the neighbor-joining algorithm and the Kimura two-parameter substitution model. RDP3 software (31) was used to test for recombination by the following four methods: an exact nonparametric method (7), sister scanning (18), the maximum χ^2 method (32), and a likelihood-assisted recombination detection method (23).

Nucleotide sequence accession numbers

The sequences of the QRDRs of the isolates studied here were submitted to GenBank under accession numbers GU002022 to GU002045.

RESULTS

Characterization of the non-levofloxacin-susceptible population

Initial screening by disk diffusion identified 47 isolates that were not susceptible to levofloxacin (14.9%), and the MICs for these isolates were determined by Etest. The MICs determined were not in full agreement with the results of the disk diffusion test. Discrepancies were observed for one isolate that was resistant by disk diffusion and that had an MIC of 4 μg/mL (intermediate by use of the CLSI criteria) and for an additional set of five isolates that were intermediate by the disk diffusion test and that had MICs of 2 μg/mL (susceptible by use of the CLSI criteria). The results of MIC testing were considered definitive for analysis in this work.

A total of 38 isolates (12.1%) were resistant to levofloxacin (MIC_{50}, >32 μg/mL; MIC range, 6 to >32 μg/mL) and 4 isolates (1.3%) expressed intermediate resistance to this antibiotic (MIC range, 3 to 4 μg/mL), resulting in a proportion of nonsusceptible isolates (MICs > 2 μg/mL) of 13.3%. Among the 30 susceptible isolates chosen for sequence analysis, the MIC range was 0.19 to 2 μg/mL.

Levofloxacin-resistant isolates were recovered from skin and soft tissue infections (*n* = 33), blood (*n* = 4), and urine (*n* = 1), while none was found among the 106 isolates recovered from the respiratory tract ($P < 0.0001$, Fisher's exact test). The recovery of small numbers of isolates in the first years of the study precluded the calculation of annual resistance rates. However, if we consider the data for the time period between 2003 and 2005, when more than 50 isolates were recovered each year, the rates of levofloxacin resistance were 11.4% in 2003 and 8.0% in 2004, and the rate reached its highest level (20.2%) in the last year of the study.

No difference in levofloxacin or ciprofloxacin MICs was observed for the nonsusceptible isolates in the presence of reserpine, excluding a role for active efflux in resistance in SDE.

Analysis of the PFGE patterns generated after digestion with SmaI of the 314 isolates revealed 12 clusters and 10 unique profiles (SID ± 95% CI, 0.817 ± 0.021). More than 80% of the isolates were distributed in four large clusters: clusters H_{88}, J_{77}, C_{56}, and E_{35} (Table 4.2). Levofloxacin-resistant isolates were present in nine of these PFGE clusters, while two isolates had unique profiles (SID ± 95% CI, 0.818 ± 0.101). PFGE cluster I_{12}

was the only cluster with more than 10 isolates that did not include any levofloxacin-resistant isolates. Of the 31 distinct *emm* types that were found in the overall population (SID ± 95% CI, 0.924 ± 0.009), 10 were represented by levofloxacin-resistant isolates (SID ± 95% CI, 0.844 ± 0.071) (Table 4.2). *stG166b* (*n* = 13) and *stG6792* (*n* = 8) accounted for more than half of the nonsusceptible isolates, followed by *stG6* (*n* = 6) and *stGLP3* (*n* = 4). *stG166b* was distributed into two main clusters: one comprised only levofloxacin-resistant isolates (grouped in cluster C_{56} [*n* = 11]) and the other mostly comprised susceptible ones (grouped in cluster J_{77} [*n* = 10], which included an isolate with intermediate susceptibility to levofloxacin) (Table 4.2). *stGLP3* was the only *emm* type exclusively found in nonsusceptible isolates and, together with the *stG166b* resistant isolates, constituted the most significant presence of levofloxacin-resistant isolates in a single PFGE cluster (PFGE cluster C_{56}; Table 4.2). The other resistant isolates were found to be dispersed in the population and were not concentrated in a specific PFGE cluster, as illustrated by *stG6792* and *stG6*.

TABLE 4.2. *emm* types and distribution of non-levofloxacin-susceptible isolates among PFGE clusters

***emm* type**	**No. of nonsusceptible isolates/no. of susceptible isolates in PFGE cluster[a]:**										
	C_{56}	**H_{88}**	**J_{77}**	**B_{11}**	**F_4**	**E_{35}**	**K_8**	**L_3**	**D_2**	**Other[b]**	**Total**
stG166b	11/0	0/1	1[d]/9						1/0		13/10
stG6792	0/2	3/28			3/0	0/1	1/0			1/1	8/32
stG6	1[d]/2	1/4	2[d]/20	1/1		1/2				0/4	6[e]/33
stGLP3	4/0										4/0
stG485	0/1	0/2	1[d]/15			0/6	0/6	1/0		1/1	3/31
stG2078	0/3		0/3	2/6				0/1			2/13
stG245			2/3			0/1			0/1	0/2	2/7
stG480	0/2	0/2	1/0	1/0		0/15		0/1		0/4	2/24
stG643		0/1	0/3			1/0				0/5	1/9
stC6979					0/1		1/0			0/3	1/4
Other[c]	0/30	0/46	0/17			0/8				0/8	0/109
Total	16/40	4/84	7/70	4/7	3/1	2/33	2/6	1/2	1/1	2/28	42/272

[a] PFGE clusters were arbitrarily designated by the use of capital letters and a subscript number, which indicates the number of isolates included in the cluster.

[b] Other included 3 PFGE clusters that did not contain resistant isolates and 10 unique PFGE profiles (i.e., the sharing of less than 80% similarity in the SmaI PFGE profile with that of any other isolate in the collection). This is the case for the two resistant isolates listed in this column.

[c] Other included 21 *emm* types that were not found in resistant isolates. Among these, only *stC839* (*n* = 30), *stG10* (*n* = 27), and *stG62647* (*n* = 11) comprised more than 10 isolates.

[d] One isolate presenting intermediate resistance is included.

[e] Two isolates presenting intermediate resistance are included.

Phylogenetic analysis of *gyrA* and *parC*

The *gyrA* and *parC* QRDR sequences were determined for 56 SDE isolates, including 22 isolates that were resistant to levofloxacin (representatives of 10 *emm* types and 11 PFGE clusters), the 4 intermediate isolates, and 30 susceptible isolates, including isolates that had the same *emm* types and that grouped together with resistant isolates in the same PFGE clusters.

The *gyrA* sequences of the SDE isolates exhibited minimal variation. Only 4 nucleotide changes were observed across the 399 bp sequenced, resulting in five distinct alleles. As these changes occurred at three specific codons and all resulted in amino acid replacements, they are all represented in Table 4.3.

TABLE 4.3. Amino acid substitutions in ParC and GyrA of SDE isolates with different levels of resistance to levofloxacin[a]

Alleles[b]	Amino acid coding change at the indicated position								MIC range (µg/mL)	No. of isolates
	GyrA			ParC						
	81	85	142	79	83	91	140	166		
gyrA04, *parC*[c]	S81F			S79F					>32	5
gyrA04, *parC15*	S81F			S79F		N91D			>32	5
gyrA04, *parC17*	S81F			S79F		N91D	P140S		>32	4
gyrA02, *parC17*	S81Y			S79F		N91D	P140S		>32	1
gyrA03, *parC08*		E85K		S79F					>32	1
gyrA03, *parC06*		E85K		S79Y					>32	1
gyrA04, *parC09*	S81F				D83N				8, 24, >32	3
gyrA04, *parC13*	S81F			S79A					12	1
gyrA03, *parC10*		E85K			D83G				8	1
gyrA04, *parC14*	S81F					N91D			4	1
gyrA01, *parC04*				S79F					3	1
gyrA01, *parC17*				S79F		N91D	P140S		1.5-3	3
gyrA01, *parC19*				S79A		N91D	P140S		2	1
gyrA01, *parC12*								G166E	1	1
gyrA01, *parC03*				S79P					0.75	1
gyrA01, *parC*[d]						N91D	P140S		0.19-0.5	5
gyrA01, *parC14*						N91D			0.5-0.75	3
gyrA05, *parC05*			D142N						0.38-0.5	4
gyrA01, *parC*[e]									0.38-0.75	14

[a] A total of 56 isolates with different levels of resistance to levofloxacin were tested.

[b] The allele numbers were attributed on the basis of the DNA sequence of the QRDR of each gene. When more than one allele that resulted in the same amino acid sequence was found, no number is indicated and the alleles found and the number of isolates presenting them are indicated in the subsequent footnotes.

[c] For *parC04*, *n* = 3; for *parC08*, *n* = 2.

[d] For *parC16*, *n* = 4; for *parC18*, *n* = 1.

[e] For *parC01*, *n* = 5; for *parC02*, *n* = 3; for *parC07*, *n* = 5; for *parC11*, *n* = 1.

In contrast, multiple point mutations (mostly silent mutations) could be observed in the *parC* sequences of both susceptible and resistant isolates, resulting in 19 distinct alleles, which are indicated in Table 4.4. Surprisingly, the *parC* sequences of the majority of the isolates analyzed (33/56, *parC01* to *parC13* alleles) were more similar to the published sequence of *S. dysgalactiae* subsp. *dysgalactiae* (GenBank accession number AB101472) than to the majority of the SDE *parC* sequences available to date in the GenBank database (GenBank accession numbers AB101473 and EF619597 to EF619601) (Fig. 4.1), with the exception of the sequence retrieved from the recently released genomic project (GenBank accession number AP010935).

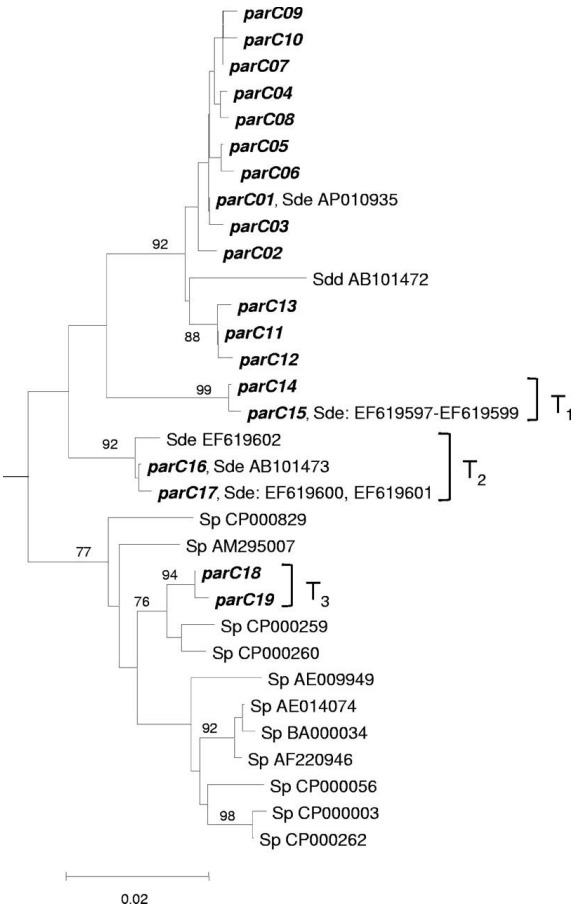

FIGURE 4.1. Neighbor-joining tree for the *parC* QRDR

The *Streptococcus agalactiae* type strain NCTC 8181 *parC* sequence (GenBank accession number AB101464) was used to root the tree. If the percentage of replicate trees in which the associated sequences clustered together in the bootstrap test (1,000 replicates) was greater than 75%, the values are shown next to the branches. The tree is drawn to scale, with the branch lengths being in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed by the Kimura two-parameter method and are in the units of the number of base substitutions per site. Sp, *Streptococcus pyogenes*; Sde, *Streptococcus dysgalactiae* subsp. *equisimilis*; Sdd, *Streptococcus dysgalactiae* subsp. *dysgalactiae*. Sets of numbers and letters identify the GenBank entries whose sequences are identical to the sequences of the alleles found in our collection or the sequences of GAS used to construct the tree.

TABLE 4.4. *parC* alleles of *S. dysgalactiae* subsp. *equisimilis* isolates

Allele	No. of isolates[a]	Base at nucleotide position (amino acid no.)[b]:																																		
		51 (17)	54 (18)	60 (20)	69 (23)	78 (26)	100 (34)	108 (36)	111 (37)	117 (39)	141 (47)	156 (52)	207 (69)	225 (75)	235 (79)	236 (79)	237 (79)	247 (83)	248 (83)	271 (91)	279 (93)	330 (110)	339 (113)	405 (135)	417 (139)	418 (140)	420 (140)	426 (142)	432 (144)	456 (152)	459 (153)	497 (166)	513 (171)	519 (173)	540 (180)	552 (184)
parC01	5	T	A	C	C	A	A	G	A	T	A	T	T	A	T	C	C	G	A	A	A	T	T	A	T	C	C	T	C	G	G	G	C	T	A	T
parC02	3										G						T																			
parC03	1														C*																					
parC04	4															T*																				
parC05	4																													A						
parC06	1															A*														A						
parC07	5																																A			
parC08	3															T*																	A			
parC09	3																	A*															A			
parC10	1																		G*														A			
parC11	1						C				G	C	C																							
parC12	1						C				G	C	C																			A*				
parC13	1						C				G	C	C		G*																					
parC14	4	C	C	T	T	G	C	A	G	C	G		C	G						G*	G	C		G												
parC15	5	C	C	T	T	G	C	A	G	C	G		C	G		T*				G*	G	C		G												
parC16	4																			G*	G	C	C	G	C	T*	T*	G		C	C			G	A
parC17	8															T*				G*	G	C	C	G	C	T*	T*	G		C	C			G	A
parC18	1	C	C	T	T	G	C	A	G	C	G		C	G						G*	G	C		G	C	T*	T*	G	T	C	C		C	G	A
parC19	1	C	C	T	T	G	C	A	G	C	G		C	G	G*					G*	G	C		G	C	T*	T*	G	T	C	C		C	G	A

[a] A total of 56 isolates with different levels of resistance to levofloxacin were tested.

[b] Base substitutions that result in amino acid coding changes are indicated by asterisks. Nucleotide and amino acid positions are numbered according to the *S. pneumoniae* numbering system (36)

The *parC18* and *parC19* alleles were most similar to the *parC* sequences of GAS strains MGAS9429 (GenBank accession number CP000259) and MGAS10270 (GenBank accession number CP000260). These alleles were found in two Lancefield group C isolates of the same *emm* type (*emm* type *stG62647*) and with identical PFGE profiles (PFGE cluster J_{77}) and differed only in the Ser-79-Ala (TCC-to-GCC) change in *parC*. The *gyrA* sequences of both isolates were indistinguishable and represented the most frequent *gyrA* allele (*gyrA01*) found in the SDE isolates in our collection, in agreement with the species identification.

Phylogenetic trees were built by the neighbor-joining method (Fig. 4.1 and 4.2). The tree for *parC* shows a single branch that gave rise to the majority of SDE *parC* alleles (*parC01* to *parC13*), which were tightly clustered together with a high bootstrap support, but also included the more divergent sequence of *S. dysgalactiae* subsp. *dysgalactiae* (Fig. 4.1). Another branch groups the more diverse GAS alleles, although with a lower level of bootstrap support. As expected, the *parC18* and *parC19* alleles (group T_3 in Fig. 4.1) of SDE were found within the GAS branch. Two additional branches supported by high bootstrap levels were identified (transformant group 1 [T_1] and T_2 in Fig. 4.1, including *parC14* to *parC17*), but they were not clearly grouped with either the SDE or the GAS branches. An analysis of the alignment with RDP3 software (31) identified two recombination events that could explain the existence of these branches. A possible recombination event (T_1) between the *parC03* and *parC18* alleles identified by an exact nonparametric method ($P = 0.0346$) and also supported by the sister-scanning method ($P = 0.0024$) could give rise to both the *parC14* and the *parC15* alleles, with the latter allele corresponding to multiple sequences deposited in the GenBank database (GenBank accession numbers EF619597 to EF619599). The *parC14* and *parC15* alleles differed at a single nucleotide, with one being more prevalent among resistant isolates (*parC15*, S79F amino acid substitution) and the other being more prevalent among susceptible isolates (Table 4.4). Another putative recombination event (T_2) was detected by the maximum χ^2 method ($P = 0.0299$) and was confirmed by the likelihood-assisted recombination detection method ($P = 0.0400$) between an GAS allele (GenBank accession number CP000260) and *parC09*. This could explain the origins of both *parC16* and *parC17*. Such a recombination event could also explain the *parC* sequences of a number of SDE strains currently deposited in the GenBank database (GenBank accession numbers EF619600 to EF619602 and AB101473).

Similar to the previously presented putative recombination event, the *parC16* allele was not found among resistant isolates, while the *parC17* allele was found to be associated with resistance (S79F amino acid substitution; see below).

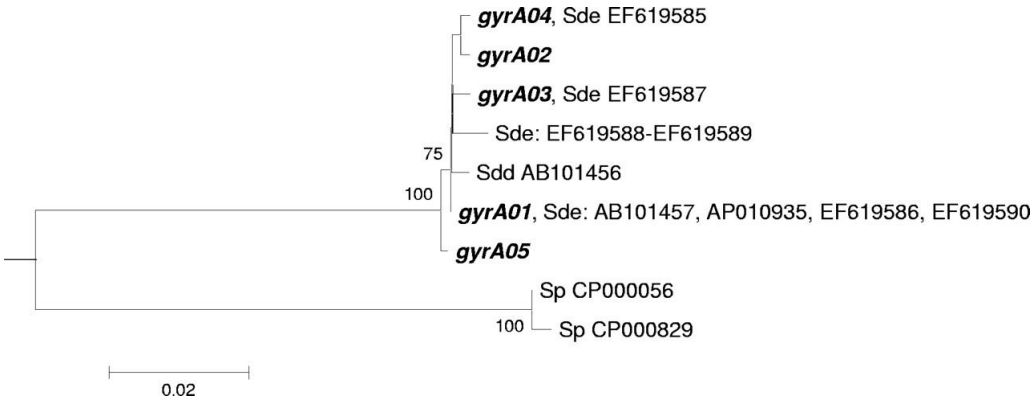

FIGURE 4.2. Neighbor-joining tree for the *gyrA* QRDR

The *Streptococcus agalactiae* type strain NCTC 8181 *gyrA* sequence (GenBank accession number AB101448) was used to root the tree. For details on the methods used to construct the tree, see the legend to Fig. 1. Sp, *Streptococcus pyogenes*; Sde, *Streptococcus dysgalactiae* subsp. *equisimilis*; Sdd, *Streptococcus dysgalactiae* subsp. *dysgalactiae*. Sets of numbers and letters identify the GenBank entries whose sequences are identical to the sequences of the alleles found in our collection or the sequences of GAS used to construct the tree.

In the phylogenetic tree for *gyrA*, all SDE alleles clustered together with high bootstrap support (Fig. 4.2). The only *gyrA* sequence of *S. dysgalactiae* subsp. *dysgalactiae* available is included in this branch and represents more than 99% sequence identity to the SDE alleles. In contrast to the case for *parC*, the GAS *gyrA* alleles are more distantly related to those of SDE.

Contribution of mutations to levofloxacin resistance

Not all the amino acid substitutions observed in the regions of *gyrA* and *parC* sequenced (4 and 9 amino acid changes, respectively) were exclusively detected in resistant isolates, as summarized in Table 4.3.

All the resistant isolates had amino acid changes within the QRDRs of both *gyrA* and *parC* compared to the consensus sequence of the susceptible isolates. The Ser-81 change to Phe (TCT to TTT) in GyrA and the Ser-79 change to Phe (TCC to TTC) in ParC were the most common combinations observed in resistant isolates (n = 14). These could also be accompanied by other changes in ParC and invariably resulted in an MIC of >32 μg/mL. These two amino acid changes were also the most widespread in

nonsusceptible isolates, and both appeared in combination with other amino acid changes (Table 4.3). Ser-81-Tyr and Glu-85-Lys in GyrA and Ser-79-Tyr, Ser-79-Ala, Asp-83-Asn, and Asp-83-Gly in ParC were other amino acid substitutions found among resistant isolates and thought to contribute to resistance. The level of levofloxacin resistance achieved was dependent on the combination of the GyrA and ParC amino acid changes present in each isolate (Table 4.3). A single set of three isolates that shared the Ser-81-Phe and Asp-83-Asn substitutions in GyrA and ParC, respectively, displayed significantly variable MIC results (range, 8 to >32 μg/mL).

In contrast to the double mutations described above, a single amino acid change in ParC resulted in intermediate resistance to levofloxacin or in slightly increased MICs in susceptible strains (0.5 to 3 μg/mL). An intermediately resistant isolate with an MIC of 4 μg/mL was also the only one to have a mutation in the GyrA QRDR but no mutations in ParC. Common to both susceptible and resistant isolates were the Asn-91-Asp (AAT to GAT) and Pro-140-Ser (CCC to TCT) substitutions in ParC, present in 23 and 14 isolates, respectively.

DISCUSSION

A surprising finding in our work was the high overall rate of resistance to levofloxacin (12%). The only previous study that specifically addressed this problem (5) found that the rate of resistance to levofloxacin was less than 1% among GGS and GCS in a collection of isolates recovered in both Europe and North America. Portugal is the European country where FQs are more widely used (19), and their sustained use has been implicated in selection for resistance (24). Differences in consumption could thus be a major factor conditioning geographic differences in FQ resistance.

Several mechanisms have been implicated in bacterial resistance to quinolones. The existence of plasmids carrying specific genes that confer increased resistance to this class of antimicrobials was recently described in the *Enterobacteriaceae* family (39). However, this resistance mechanism has thus far not been found in Gram-positive cocci, in which the increased expression of multidrug resistance pumps and alterations of the fluoroquinolone targets are the major mechanisms of quinolone resistance (24, 30). Similar to what was described in other streptococci, *in vitro* studies with GAS, a closely related species, have implicated alterations in the QRDRs of *gyrA* and *parC* as the primary mechanism of the loss of susceptibility to fluoroquinolones (6). Studies with *Streptococcus pneumoniae* showed that the transformation of a susceptible strain with the QRDRs of these genes from a resistant strain was sufficient to confer resistance (43). Taken together, these findings indicate that alterations of the QRDRs of *gyrA* and *parC* are necessary and sufficient for the development of quinolone resistance in streptococci. We found no evidence of efflux-mediated resistance to levofloxacin in SDE, similar to what was described in GAS (30, 38), indicating that this resistance mechanism is not widespread in beta-hemolytic streptococci and leaving the alterations in the quinolone targets as the sole and most probable mechanism of resistance.

The levofloxacin-resistant isolates presented a variety of *emm* types, comparable to that observed for the entire collection, and they were distributed in several PFGE clusters, indicating a polyclonal origin of resistance. It is known from epidemiological studies of GAS that some resistant lineages predominate, although mutational alterations can also be observed in diverse genetic backgrounds (1, 29). For instance, the ciprofloxacin-resistant lineage of GAS carrying the *emm6* allele is disseminated in distinct geographic regions, such as Spain (1), Belgium (29, 30), and the United States

(35). In our collection, we could also identify a PFGE cluster (PFGE cluster C_{56}) that was significantly associated with resistance to levofloxacin (P = 0.0009, Fisher's exact test), and only two of the *emm* types found in this cluster (*emm* types *stG166b* and *stGLP3*) presented resistant isolates. This suggests that stable levofloxacin-resistant lineages also exist in SDE. However, the latter did not dominate among the levofloxacin-resistant SDE isolates recovered in Portugal. When we tested if the high rate of resistance found in the last year of the study could have been caused by an outbreak of a single clone, we verified that this was not so, since the isolates recovered in 2005 were a heterogeneous population (seven *emm* types and nine PFGE clusters).

Among the streptococci, horizontal gene transfer involving the QRDRs of the FQ target genes was first described for *S. pneumoniae* and members of the viridans group streptococci (3, 16). More recently, the *parC* QRDRs of GAS were also shown to present sequence polymorphisms (35), and these observations have been the basis for the suggestion of the occurrence of horizontal gene transfer events in this species (14, 38), with *S. dysgalactiae* acting as the donor. *S. dysgalactiae* subsp. *dysgalactiae* was suggested to be the most likely source of DNA by Duesberg *et al.*. (14). However, such conclusions have been hampered by the lack of SDE *parC* sequences, as only one *parC* sequence of an SDE isolate was publicly available and was found in the GenBank database (GenBank accession number AB101473) at the time of these studies. *S. dysgalactiae* subsp. *dysgalactiae* is mainly a pathogen of animals and, contrary to the case for SDE, does not usually infect humans (8, 15). No other reservoir for GAS besides the human host is known, making it unlikely that situations favoring gene exchange with *S. dysgalactiae* subsp. *dysgalactiae* would arise. Thus, SDE would seem to be a more suitable candidate that engages in genetic exchange with GAS, as they can share ecological niches in the human host. Previous studies have documented the existence of bacteriophages capable of the *in vitro* transduction of chromosomal markers from GAS to GCS (12), as well as from GCS to GAS (44), providing a mechanism that could mediate genetic exchange between these two species. In agreement with the findings of previous studies, our results support the acquisition of GAS DNA by SDE (25). The SDE *parC18* and *parC19* alleles cluster within the GAS branch, but the *gyrA* alleles of the isolates in which the *parC* alleles were found were characteristic of those of SDE. Taken together, these data strongly suggest that substitution of the entire QRDR of *parC* occurred in these isolates, through

recombination with GAS DNA. The *parC14* to *parC17* alleles resulted from recombination within the QRDR with GAS alleles, with DNA stretches characteristic of both species being retained (Fig. 4.3). Interestingly, the QRDR of the *parC* sequence first available in the GenBank database for SDE seems to have also resulted from recombination, a fact that may have confounded previous analyses that led to the suggestion that *S. dysgalactiae* subsp. *dysgalactiae* could have been the donor species for recombination with GAS. Our data support the ecologically more plausible hypothesis that gene exchange between SDE and GAS is more frequent than that between *S. dysgalactiae* subsp. *dysgalactiae* and GAS. Horizontal transfer within SDE may also be ongoing, since the same QRDR polymorphisms were found among distinct lineages, i.e., isolates belonging to different PFGE clusters and with distinct *emm* types (data not shown). No recombination events between SDE and GAS were detected in the *gyrA* QRDR. A greater divergence between these species at this *locus* could explain this observation, since sequence divergence is a major barrier to gene exchange (28).

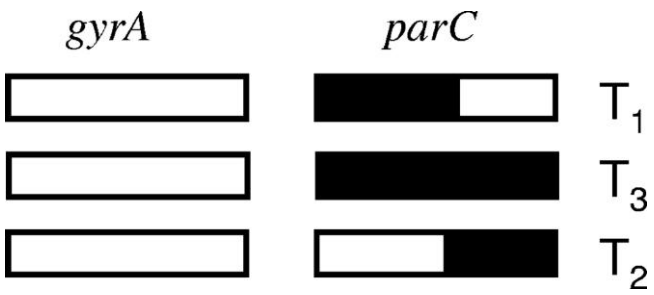

FIGURE 4.3. Diagram of the QRDRs of *gyrA* and *parC* in isolates showing recombination events

The box on the left represents the *gyrA* QRDR, while that on the right represents the *parC* QRDR. White boxes represent regions that are characteristic of SDE, while black boxes represent regions characteristic of GAS. Each allele combination represents the sequences resulting from the putative recombination events indicated in Fig. 4.1 (T_1 to T_3). Recombination breakpoints were determined by the maximum χ^2 method implemented in RDP3 software, refined by visual inspection of the sequences (Table 4.4).

All recombination events detected included both susceptible and resistant recombinants, indicating that there was not a strict relationship between recombination with foreign DNA and resistance and raising the possibility that the mutations conferring levofloxacin resistance occurred subsequent to DNA acquisition. In fact, the quinolone resistance-defining mutations have probably arisen independently in multiple strains sharing the same mosaic *parC* genes. If this scenario proves to be true, the expansion of levofloxacin resistance among SDE isolates occurs mostly through vertical

transmission and not through horizontal dissemination. Still, although our study was not designed to evaluate the frequency of recombination, we did identify a high proportion of recombinant QRDRs (*n* = 23/56 [41%] isolates), suggesting that recombination is frequent and ongoing.

Among our collection we could identify several pairs of levofloxacin-resistant and -susceptible isolates that shared the same *emm* type or SmaI profile, allowing us to investigate more closely the origin of levofloxacin resistance in these isolates. Only when mutations were concomitantly present in *gyrA* and *parC* were the isolates levofloxacin resistant. Changes to phenylalanine in Ser-81 in GyrA and Ser-79 in ParC were the most important in the acquisition of levofloxacin resistance in terms of both the frequency of occurrence and the level of resistance achieved, while mutations of only one of these amino acids predominated in isolates with reduced susceptibility. Our data support the hypothesis that the target for first-step mutations in SDE is mainly ParC, as previously observed in other streptococci (34). A few groups of isolates with the same QRDR alleles showed some MIC variations, but most values were within a 1-dilution interval.

Isolates presenting high-level resistance to fluoroquinolones were described among different species of beta-hemolytic streptococci, but only one previous work described the mutations present in SDE leading to fluoroquinolone resistance (5). The amino acid substitutions found to be present in the FQ-resistant isolates among our isolates recovered in Portugal resulted from changes in the same relative positions in GyrA (Ser-81, Glu-85) and ParC (Ser-79, Asp-83), as was the case with the changes described in other beta-hemolytic streptococcal species (5, 26). The Ser-79-Ala, Ser-79-Tyr, Asp-83-Gly, and Asp-83-Asn modifications, all in ParC, were not previously described in SDE.

Quinolone resistance in SDE was mediated by alterations in ParC and GyrA and not by efflux, similar to the situation in GAS. Sequence polymorphisms not related to FQ resistance are common in the *parC* QRDR of *S. dysgalactiae* subsp. *equisimilis*; in contrast, the QRDR sequence of *gyrA* is highly conserved. Our data question the previously proposed role of *S. dysgalactiae* subsp. *dysgalactiae* as a donor for the mosaic structures found in GAS (14) and support the more plausible explanation that the exchanges occur with SDE isolates that share ecological niches with GAS. A close proximity between strains of the two species would afford bacteriophages the opportunity to transduce the genetic information between them. The high proportion of

recombinant sequences identified supports the occurrence of frequent genetic exchanges between the two species that do not seem to be driven exclusively by antimicrobial usage.

ACKNOWLEDGMENTS

This work was partly supported by the Fundação para a Ciência e a Tecnologia (grant PTDC/SAU-ESA/72321/2006); the Fundação Calouste Gulbenkian, Portugal; and an unrestricted grant from GlaxoSmithKline, Portugal.

Members of the Portuguese Group for the Study of Streptococcal Infections are as follows: Luís Lito, and Lurdes Monteiro, Hospital de Santa Maria, Lisbon; Filomena Martins, Maria Ana Pessanha, Elsa Gonçalves, and Teresa Morais, Hospital de São Francisco Xavier, Lisbon; José Diogo, Ana Rodrigues, and Isabel Nascimento, Hospital Garcia de Orta, Almada; Manuela Ribeiro, Fernanda Cotta, and Dolores Pinheiro, Hospital de São João, Porto; Paulo Lopes, Ismália Calheiros, Luísa Felício, and Angelina Lameirão, Centro Hospitalar de Vila Nova de Gaia, Vila Nova de Gaia; Valquíria Alves, Antónia Read, and Margarida Monteiro, Hospital Pedro Hispano, Matosinhos; Ana Paula Castro, Hospital de Vila Real, Vila Real; Ana Paula M. Vieira, Hospital Senhora da Oliveira, Guimarães; Elmano Ramalheira, António Bastos Marques, and Fernanda Bessa, Hospital Infante Dom Pedro, Aveiro; Rui Tomé, Graça Ribeiro, Celeste Pontes, and Fátima Martins, Hospitais da Universidade de Coimbra, Coimbra; Helena Ramos, Ana Paula Castro, Susana Ferreira, and Paulo Pinto, Hospital de Santo António, Porto; Maria Alberta Faustino and Adelaide Alves, Hospital de São Marcos, Braga; and M. F. Menezes, Instituto Bacteriológico Câmara Pestana, Lisbon.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: M. D. Pinho, J. Melo-Cristino and M. Ramirez. Performed the experiments: M. D. Pinho. Analyzed the data: M. D. Pinho, J. Melo-Cristino and M. Ramirez. Contributed reagents/materials/analysis tools: J. Melo-Cristino and M. Ramirez. Wrote the paper: M. D. Pinho, J. Melo-Cristino and M. Ramirez **.**

REFERENCES

1. **Albertí S, Cortes G, Garcia-Rey C, Rubio C, Baquero F, Garcia-Rodriguez JA, Bouza E, Aguilar L.** 2005. *Streptococcus pyogenes* pharyngeal isolates with reduced susceptibility to ciprofloxacin in Spain: mechanisms of resistance and clonal diversity. Antimicrob. Agents Chemother. **49**:418-420.

2. **Alou L, Ramirez M, Garcia-Rey C, Prieto J, Lencastre H**. 2001. *Streptococcus pneumoniae* isolates with reduced susceptibility to ciprofloxacin in Spain: clonal diversity and appearance of ciprofloxacin-resistant epidemic clones. Antimicrob. Agents Chemother. **45**:2955-2957.

3. **Balsalobre L, Ferrandiz MJ, Linares J, Tubau F, de la Campa AG**. 2003. Viridans group streptococci are donors in horizontal transfer of topoisomerase IV genes to *Streptococcus pneumoniae*. Antimicrob. Agents Chemother. **47**:2072-2081.

4. **Bast DJ, Low DE, Duncan CL, Kilburn L, Mandell LA, Davidson RJ, de Azavedo JC**. 2000. Fluoroquinolone resistance in clinical isolates of *Streptococcus pneumoniae*: contributions of type II topoisomerase mutations and efflux to levels of resistance. Antimicrob. Agents Chemother. **44**:3049-3054.

5. **Biedenbach DJ, Toleman MA, Walsh TR, Jones RN**. 2006. Characterization of fluoroquinolone-resistant beta-hemolytic *Streptococcus* spp. isolated in North America and Europe including the first report of fluoroquinolone-resistant *Streptococcus dysgalactiae* subspecies *equisimilis*: report from the SENTRY Antimicrobial Surveillance Program (1997-2004). Diagn. Microbiol. Infect. Dis. **55**:119-127.

6. **Billal DS, Fedorko DP, Yan SS, Hotomi M, Fujihara K, Nelson N, Yamanaka N**. 2007. In vitro induction and selection of fluoroquinolone-resistant mutants of *Streptococcus pyogenes* strains with multiple *emm* types. J. Antimicrob. Chemother. **59**:28-34.

7. **Boni MF, Posada D, Feldman MW**. 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics **176**:1035-1047.

8. **Brandt CM, Spellerberg B**. 2009. Human infections due to *Streptococcus dysgalactiae* subspecies *equisimilis*. Clin. Infect. Dis. **49**:766-772.

9. **Brenwald NP, Gill MJ, Wise R**. 1998. Prevalence of a putative efflux mechanism among fluoroquinolone-resistant clinical isolates of *Streptococcus pneumoniae*. Antimicrob. Agents Chemother. **42**:2032-2035.

10. **Carriço JA, Silva-Costa C, Melo-Cristino J, Pinto FR, de Lencastre H, Almeida JS, Ramirez M**. 2006. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. J. Clin. Microbiol. **44**:2524-2532.

11. **Clinical and Laboratory Standards Institute**. 2008. Performance standards for antimicrobial susceptibility testing, 16th informational supplement. CLSI document M100-S18. Clinical and Laboratory Standards Institute, Wayne, PA.

12. **Colón AE, Cole RM, Leonard CG**. 1972. Intergroup lysis and transduction by streptococcal bacteriophages. J. Virol. **9**:551-553.

13. **De La Campa AG, Ardanuy C, Balsalobre L, Perez-Trallero E, Marimon JM, Fenoll A, J. Liñares J**. 2009. Changes in fluoroquinolone-resistant *Streptococcus pneumoniae* after 7-valent conjugate vaccination, Spain. Emerg. Infect. Dis. **15**:905-911.

14. **Duesberg CB, Malhotra-Kumar S, Goossens H, McGee L, Klugman KP, Welte T, Pletz MW**. 2008. Interspecies recombination occurs frequently in quinolone resistance-determining regions of clinical isolates of *Streptococcus pyogenes*. Antimicrob. Agents Chemother. **52**:4191-4193.

15. **Facklam R**. 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin. Microbiol. Rev. **15**:613-630.

16. **Ferrandiz MJ, Fenoll A, Linares J, De La Campa AG**. 2000. Horizontal transfer of *parC* and *gyrA* in fluoroquinolone-resistant clinical isolates of *Streptococcus pneumoniae*. Antimicrob. Agents Chemother. **44**:840-847.

17. **Figueiredo TA, Aguiar SI, Melo-Cristino J, Ramirez M**. 2006. DNA methylase activity as a marker for the presence of a family of phage-like elements conferring efflux-mediated macrolide resistance in streptococci. Antimicrob. Agents Chemother. **50**:3689-3694.

18. **Gibbs MJ, Armstrong JS, Gibbs AJ**. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics **16**:573-582.

19. **Goossens H, Ferech M, Coenen S, Stephens P**. 2007. Comparison of outpatient systemic antibacterial use in 2004 in the United States and 27 European countries. Clin. Infect. Dis. **44**:1091-1095.

20. **Guerin F, Varon E, Hoi AB, Gutmann L, Podglajen I**. 2000. Fluoroquinolone resistance associated with target mutations and active efflux in oropharyngeal colonizing isolates of viridans group streptococci. Antimicrob. Agents Chemother. **44**:2197-2200.

21. **Haidan A, Talay SR, Rohde M, Sriprakash KS, Currie BJ, Chhatwal GS**. 2000. Pharyngeal carriage of group C and group G streptococci and acute rheumatic fever in an Aboriginal population. Lancet 356:1167-1169.

22. **Hashikawa S, Iinuma Y, Furushita M, Ohkura T, Nada T, Torii K, Hasegawa T, Ohta M**. 2004. Characterization of group C and G streptococcal strains that cause streptococcal toxic shock syndrome. J. Clin. Microbiol. **42**:186-192.

23. **Holmes EC, Worobey M, Rambaut A**. 1999. Phylogenetic evidence for recombination in dengue virus. Mol. Biol. Evol. **16**:405-409.

24. **Hooper DC**. 2002. Fluoroquinolone resistance among Gram-positive cocci. Lancet Infect. Dis. **2**:530-538.

25. **Kalia A, Bessen DE**. 2003. Presence of streptococcal pyrogenic exotoxin A and C genes in human isolates of group G streptococci. FEMS Microbiol. Lett. **219**:291-295.

26. **Kawamura Y, Fujiwara H, Mishima N, Tanaka Y, Tanimoto A, Ikawa S, Itoh Y, Ezaki T**. 2003. First *Streptococcus agalactiae* isolates highly resistant to quinolones, with point mutations in gyrA and parC. Antimicrob. Agents Chemother. **47**:3605-3609.

27. **Lopardo HA, Vidal P, Sparo M, Jeric P, Centron D, Facklam RR, Paganini H, Pagniez NG, Lovgren M, Beall B**. 2005. Six-month multicenter study on invasive

infections due to *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis* in Argentina. J. Clin. Microbiol. **43**:802-807.

28. **Majewski J**. 2001. Sexual isolation in bacteria. FEMS Microbiol. Lett. **199**:161-169.

29. **Malhotra-Kumar S, Lammens C, Chapelle S, Mallentjer C, Weyler J, Goossens H**. 2005. Clonal spread of fluoroquinolone non-susceptible *Streptococcus pyogenes*. J. Antimicrob. Chemother. **55**:320-325.

30. **Malhotra-Kumar S, Van Heirstraeten L, Lammens C, Chapelle S, Goossens H**. 2009. Emergence of high-level fluoroquinolone resistance in *emm*6 *Streptococcus pyogenes* and *in vitro* resistance selection with ciprofloxacin, levofloxacin and moxifloxacin. J. Antimicrob. Chemother. **63**:886-894.

31. **Martin DP**. 2009. Recombination detection and analysis using RDP3. Methods Mol. Biol. **537**:185-205.

32. **Maynard Smith J**. 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. **34**:126-129.

33. **McDonald M, Towers RJ, Andrews RM, Carapetis JR, Currie BJ**. 2007. Epidemiology of *Streptococcus dysgalactiae* subsp. *equisimilis* in tropical communities, Northern Australia. Emerg. Infect. Dis. **13**:1694-1700.

34. **Muñoz R, De La Campa AG**. 1996. ParC subunit of DNA topoisomerase IV of *Streptococcus pneumoniae* is a primary target of fluoroquinolones and cooperates with DNA gyrase A subunit in forming resistance phenotype. Antimicrob. Agents Chemother. **40**:2252-2257.

35. **Orscheln RC, Johnson DR, Olson SM, Presti RM, Martin JM, Kaplan EL, Storch GA**. 2005. Intrinsic reduced susceptibility of serotype 6 *Streptococcus pyogenes* to fluoroquinolone antibiotics. J. Infect. Dis. **191**:1272-1279.

36. **Pan XS, Ambler J, Mehtar S, Fisher LM**. 1996. Involvement of topoisomerase IV and DNA gyrase as ciprofloxacin targets in *Streptococcus pneumoniae*. Antimicrob. Agents Chemother. **40**:2321-2326.

37. **Pinho MD, Melo-Cristino J, Ramirez M**. 2006. Clonal relationships between invasive and non-invasive Lancefield group C and G streptococci and *emm*-specific differences in invasiveness. J. Clin. Microbiol. **44**:841-846.

38. **Pletz MW, McGee L, Van Beneden CA, Petit S, Bardsley M, Barlow M, Klugman KP**. 2006. Fluoroquinolone resistance in invasive *Streptococcus pyogenes* isolates due to spontaneous mutation and horizontal gene transfer. Antimicrob. Agents Chemother. **50**:943-948.

39. **Robicsek A, Jacoby GA, Hooper DC**. 2006. The worldwide emergence of plasmid-mediated quinolone resistance. Lancet Infect. Dis. **6**:629-640.

40. **Sachse S, Seidel P, Gerlach D, Gunther E, Rodel J, Straube E, Schmidt KH**. 2002. Superantigen-like gene(s) in human pathogenic *Streptococcus dysgalactiae*, subsp. *equisimilis*: genomic localisation of the gene encoding streptococcal pyrogenic exotoxin G (*speG*^{dys}). FEMS Immunol. Med. Microbiol. **34**:159-167.

41. **Siljander T, Karppelin M, Vahakuopus S, Syrjanen J, Toropainen M, Kere J, Vuento R, Jussila T, Vuopio-Varkila J**. 2008. Acute bacterial, nonnecrotizing cellulitis in Finland: microbiological findings. Clin. Infect. Dis. **46**:855-861.

42. **Tamura K, Dudley J, Nei M, Kumar S**. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**:1596-1599.

43. **Varon E, Janoir C, Kitzis MD, Gutmann L**. 1999. ParC and GyrA may be interchangeable initial targets of some fluoroquinolones in *Streptococcus pneumoniae*. Antimicrob. Agents Chemother. **43**:302-306.

44. **Wannamaker LW, Almquist S, Skjold S**. 1973. Intergroup phage reactions and transduction between group C and group A streptococci. J. Exp. Med. **137**:1338-1353.

45. **Woo PC, Fung AM, Lau SK, Wong SS, Yuen KY**. 2001. Group G beta-hemolytic streptococcal bacteremia characterized by 16S ribosomal RNA gene sequencing. J. Clin. Microbiol. **39**:3147-3155.

CHAPTER 5

Multilocus sequence analysis of *Streptococcus canis* confirms the zoonotic origin of human infections and reveals genetic exchange with *Streptococcus dysgalactiae* subsp. *equisimilis*

The contents of this chapter were presented in:

Pinho MD, Matos SC, Pomba C, Lübke-Becker A, Wieler LH, Preziuso S, Melo-Cristino J, Ramirez M

2013. Journal of Clinical Microbiology **51**: 1099-1109

ABSTRACT

Streptococcus canis is an animal pathogen that occasionally causes human infections. Isolates recovered from infections of animals (n = 78, recovered from 2000 to 2010 in three European countries, mainly from house pets) and humans (n = 7, recovered from 2006 to 2010 in Portugal) were identified by phenotypic and genotypic methods and characterized by antimicrobial susceptibility testing, multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and *emm* typing. *S. canis* isolates presented considerable variability in biochemical profiles and 16S rRNA. Resistance to antimicrobial agents was low, with the most significant being *tet*(M)- and *tet*(O)-mediated tetracycline resistance. MLST analysis revealed a polyclonal structure of the *S. canis* population causing infections, where the same genetic lineages were found infecting house pets and humans and were disseminated in distinct geographic locations. Phylogenetic analysis indicated that *S. canis* was a divergent *taxon* of the sister species *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis* (SDE) and found evidence of acquisition of genetic material by *S. canis* from SDE. PFGE confirmed the MLST findings, further strengthening the similarity between animal and human isolates. The presence of *emm*-like genes was restricted to a few isolates and correlated with some MLST-based genetic lineages, but none of the human isolates could be *emm* typed. Our data show that *S. canis* isolates recovered from house pets and humans constitute a single population and demonstrate that isolates belonging to the main genetic lineages identified have the ability to infect the human host, providing strong evidence for the zoonotic nature of *S. canis* infection.

INTRODUCTION

Streptococcus canis is a Lancefield group G beta-hemolytic streptococcal species which is mainly found as an animal colonizer and pathogen. This member of the large-colony-forming Lancefield group C and G streptococcus (GCGS) group (1) was officially established as a distinct *taxon* in 1986 (2), following previous studies which showed that the Lancefield group G streptococci isolated from animals and humans were biochemically divergent and could represent distinct species (3). Although *S. canis* can be part of the female reproductive tract and the tonsillar and ear microbiota of cats and dogs (4–7), it is also an important pathogen for these two species and infects a wide range of other domestic and wild animals (2, 8–10). *S. canis* is the most common streptococcal species found in dog infections (11), being identified in cases of dermatitis, otitis externa, pneumonia, infective endocarditis, and adult septicemia (6, 11, 12). This species has also been implicated in fetal or neonatal septicemia, leading to abortion or neonatal death, respectively (2, 11), and in both canine and feline necrotizing fasciitis and streptococcal toxic shock syndrome (13, 14). *S. canis* may be transmitted between different animal species living in proximity (15) and has been responsible for outbreaks of clinical and subclinical mastitis with bacterial shedding in milk in cattle herds (16, 17) and in pets living in shelters (14). However, little is known about the genetic diversity of *S. canis*, namely, if there are clones particularly adapted to the different animal species.

The prevalence of *S. canis* in human infections remains unclear. The first confirmed report of human infection was published in 1997 and consisted of a case of septicemia in a 77-year-old man (18). In the last few years, a better awareness of *S. canis* has led to an increasing number of reports in the literature (19–23). *S. canis* has mainly been isolated from cases of bacteremia (18, 19, 23) and skin and soft tissue infections (20, 21). Some of these studies indicated that contact of pets with patients' wounds and ulcers would be the most likely portal of entry (18, 21), and one report identified the same *S. canis* strain in a dog and its owner's blood cultures following a bite (22). These observations, together with its wide occurrence in animal hosts, led to the perception that *S. canis* is an occasional zoonotic agent, while the majority of Lancefield group G streptococcal infections in humans are due to *Streptococcus dysgalactiae* subsp. *equisimilis* (SDE).

S. canis isolates can be distinguished from group G SDE by biochemical and molecular tests (23–25). The analysis of 16S and 23S rRNA sequences and that of other housekeeping genes also allows the distinction between these two taxa (17, 23) and shows a close genetic relationship of both to another beta-hemolytic streptococcal species, namely, *Streptococcus pyogenes* (Lancefield group A streptococcus [GAS]) (26). Inter- and intraspecies differentiation of *S. canis* isolates may also be achieved by using various typing methods. Macrorestriction profiling using pulsed-field gel electrophoresis (PFGE) has been extensively used to characterize animal and human streptococci, including *S. canis* (13, 22, 27). *emm* typing, a technique relying on sequence analysis of the M-protein gene developed for GAS and also applicable to SDE (http://www.cdc.gov/ncidod/biotech/strep/M-ProteinGene_typing.htm), has also been employed (19, 23). More recently, a multilocus sequence typing (MLST) scheme has been proposed to characterize both SDE and *S. canis*, but it was unsuccessful in typing all *S. canis* isolates (26).

The aim of the current work was to characterize a collection of *S. canis* isolates recovered from animals (mainly house pets) and humans. The molecular typing of these isolates allowed us to define the clonal structure of the *S. canis* population and to directly compare isolates from different hosts. The current study describes the first MLST analysis of *S. canis* from nonhuman sources and expands considerably the known diversity of MLST alleles and sequence types (STs) for this species.

MATERIALS AND METHODS

Bacterial isolates

A total of 85 *S. canis* clinical isolates which were recovered from house pets (*n* = 77), humans (*n* = 7), and a horse (*n* = 1) were studied (Table 5.1). Animal isolates were recovered at the Faculty of Veterinary Medicine of the Technical University of Lisbon, Lisbon, Portugal (*n* = 48, isolated between 2001 and 2007), the Institute of Microbiology and Epizootics, Veterinary Faculty, Freie Universität Berlin, Berlin, Germany (*n* = 29, isolated between 2000 and 2010), and the University of Camerino, Matelica, Italy (*n* = 1, isolated in 2007). Companion animals were individually owned, and no epidemiologic relationship was evident upon admission to the collecting laboratory. The seven human isolates were collected during an ongoing nationwide study of human large-colony-forming GCGS infections in Portugal between 2006 and 2010 in six distinct hospital laboratories. No information on patient contact with animals was recorded. The *S. canis* type strain DSM 20715, isolated from a case of bovine mastitis (2), was also analyzed by MLST, PFGE, and *emm* typing for comparison purposes.

TABLE 5.1. Distribution of the 85 *S. canis* clinical isolates studied by source, origin, and geographic location

Source	No. of isolates by origin and geographic location[a]												Total
	Dog			Cat			Horse			Human			
	P	G	I	P	G	I	P	G	I	P	G	I	
Ear exudate	27	1		1									29
Skin and soft tissue exudate	5	4		1	6					1			17
Urine	14												14
Vaginal exudate		8											8
Blood		2								4			6
Respiratory tract secretions		2								2			4
Eye exudate		3											3
Tonsillar exudate		1	1										2
Gastrointestinal sample								1					1
Unknown		1											1
Total	46	22	1	2	6			1		7			85

[a] P, Portugal; G, Germany; I, Italy.

Species identification

The Lancefield group was confirmed by a commercial latex agglutination technique (Slidex Strepto kit; bioMérieux, Marcy l'Etoile, France). Beta-hemolysis and colony size were confirmed in tryptic soy agar (Oxoid, Hampshire, England) supplemented with 5% (vol/vol) defibrinated sheep blood, after overnight incubation at 37°C.

Species identification was obtained by biochemical and genotypic methods. All isolates were identified by using the API 20 Strep system (bioMérieux, Marcy l'Etoile, France) and by applying a PCR targeting 23S rRNA to differentiate *S. canis* from SDE (25), which can also present the Lancefield group G antigen. Briefly, primers dysF (5′-GCGACATGGGACTTTAAAAGG-3′) and canR (5′-CTCCTACCATTACCTCTTAAGGTA-3′) were used to amplify a 924-bp fragment of the *S. canis* 23S rRNA gene. PCR conditions were identical to those previously described (25).

The 16S rRNA of over one-third of the *S. canis* isolates in our collection was amplified and sequenced with previously described primers and conditions (28). The group selected included isolates for which the API profiles presented a confidence of identification below 99% (according to the manufacturer's literature) and isolates representing different MLST groups.

Antimicrobial susceptibility testing and genotype determination

Testing for susceptibility to penicillin, cefotaxime, erythromycin, clindamycin, tetracycline, chloramphenicol, levofloxacin, vancomycin, quinupristin-dalfopristin, and linezolid was performed by disk diffusion according to CLSI guidelines for beta-hemolytic streptococci (29). High-level resistance to gentamicin and streptomycin was screened by using the CLSI recommendations for *Enterococcus* spp. (29). MICs for macrolide- and aminoglycoside-resistant isolates were determined using Etest strips (bioMérieux, Marcy l'Etoile, France) according to the manufacturer's instructions and CLSI interpretative criteria (29). A multiplex PCR scheme was used for the amplification of *erm*(A) [subclass *erm*(TR)], *erm*(B), and *mef*(A/B), conferring resistance to macrolides (30). The tetracycline resistance determinants *tet*(M), *tet*(O), *tet*(L), and *tet*(K) were detected using a multiplex PCR scheme (31). Primers for the *tet*(S) gene (32) were also included. The *aadA* and *aadE* genes, conferring resistance to

streptomycin but not to gentamicin, were amplified using the primers described by Clark *et al.* (33).

MLST

S. canis isolates were tested with a set of seven primer pairs previously described for the MLST characterization of this species and SDE (26). We found that most *S. canis* isolates failed to yield a PCR product for the *xpt* gene using these primers, in accordance with the previous observation that an *xpt* gene could not be amplified from all *S. canis* isolates (26). Primers Xptgc-up (5′-TTACTTGAAGAACGCATCTTA-3′) and Xptgc-dn (5′-ATGAGGTCACTTCAATGCCC-3′), designed for SDE in a distinct study (34), were used instead for amplification and sequencing of this *locus*. The sequences obtained were trimmed to cover the same region of the *xpt* gene reported previously (26), and alleles identical to those described earlier were determined for some isolates, indicating that the same *locus* is targeted by using both primer sets, as observed for SDE (34).

Unique sequences at each *locus* were assigned allele numbers. The combination of the seven allele numbers for each isolate was used to define the ST. ST assignment was done following the numbering started by Ahmad *et al.*. for human invasive isolates recovered in the United States (26). Data for three of these isolates were included in the analysis (corresponding to ST2 to ST4). The ST1 isolate in the North American study failed to yield a PCR product for the *xpt* gene, so it was excluded from the current analysis. ST1 in our study was assigned to the ST that shared all other six alleles with that isolate. The MLST data for *S. canis* strain FSL Z3-227 were extracted from the genome shotgun sequence assemblies obtained from GenBank (accession numbers AIDX01000001 to AIDX01000003) (35) and also included in the analysis. Strain FSL Z3-227 was isolated in an outbreak of bovine subclinical mastitis in the United States (15). The goeBURST algorithm (36) implemented in PHYLOViZ software (37) was used to establish relationships between STs. Clonal complexes (CCs) were defined at the single-locus-variant (SLV) and double-locus-variant (DLV) levels.

Phylogenetic analysis of the 16S rRNA and MLST *loci*

Sequence alignments were performed manually. The MEGA (version 5) program (38) was used to construct trees by using the neighbor-joining algorithm and the Kimura

two-parameter substitution model. Branch support was tested by 1,000 replicate bootstrap tests in each analysis. The 16S rRNA gene sequences of the type strains *S. canis* DSM 20715 (GenBank accession number DQ303184), SDE CIP105120 (GenBank accession number DQ232540), *S. dysgalactiae* subsp. *dysgalactiae* ATCC 43078 (GenBank accession number AB002485), *S. agalactiae* ATCC 13813 (GenBank accession number DQ303183), and *S. suis* ATCC 43765 (GenBank accession number DQ303193) were used for comparison. The closest match to the 16S rRNA gene of the GAS type strain allowing comparison to the 1,435 bp analyzed was selected for construction of the tree (GenBank accession number CP003068). MLST sequence data from GAS and SDE used for the phylogenetic comparison with *S. canis loci* were obtained from the respective MLST databases (http://www.mlst.net/). The *yqiZ* (also called *atoB*) (34) gene used in the SDE MLST scheme and the *yqiL* gene used for GAS are distinct *loci* and were excluded when concatenating the internal fragments of the genes used in MLST.

PFGE and *emm* typing

Isolates were characterized by PFGE macrorestriction profiling and by *emm* typing as previously described (28). Briefly, PFGE patterns generated after SmaI (Fermentas, Vilnius, Lithuania) digestion were compared by using Bionumerics software (Applied Maths, Sint-Martens-Latem, Belgium) to create dendrograms by the unweighted-pair group method with arithmetic means (UPGMA). The Dice similarity coefficient was used with optimization and position tolerance settings of 1.0 and 1.5, respectively. Clusters were defined as groups of isolates ($n \geq 2$) presenting profiles ≥80% related on the dendrogram (28). Characterization by *emm* typing was carried out with the primers and conditions available at http://www.cdc.gov/ncidod/biotech/strep/protocol_emm-type.htm.

Statistical analysis

The diversity of the classifications and the congruence of the results obtained by PFGE and MLST were quantitatively evaluated by calculating Simpson's index of diversity (SID) (39) and the adjusted Wallace (AW) coefficient (40) with 95% confidence intervals (CIs). All calculations were done using the Comparing Partitions website (http://darwin.phyloviz.net/ComparingPartitions/).

Nucleotide sequence accession numbers

The sequences of the 16S rRNA and MLST alleles found in the isolates studied here were submitted to GenBank under accession numbers JX876612 to JX876641. The sequence of the novel *emm* type *stG1451* was submitted to GenBank (accession number KC291148) and to the CDC database (http://www.cdc.gov/ncidod/biotech/strep/M-ProteinGene_typing.htm). An online database using the BIGSdb platform (40a) and containing the data reported in this paper is available at http://pubmlst.org/scanis/.

RESULTS

Phenotypic and genotypic identification of *S. canis* isolates

All isolates were beta-hemolytic and carried the Lancefield group G polysaccharide. The API 20 Strep system identified all but one isolate as *S. canis*. Among the 26 distinct biochemical profiles that were found with this system, 21 (n = 81 isolates, including the type strain DSM 20715) corresponded to *S. canis* with a confidence of ≥99% according to the manufacturer's literature, while for 4 isolates it ranged from 87.6% to 97.6%. Only one isolate had a profile not consentaneous with *S. canis*, being identified as group L streptococcus with 86.0% confidence (API 20 Strep profile 2673415).

All isolates yielded a PCR product of the expected size for *S. canis* in the 23S rRNA PCR assay, while an SDE control isolate was negative (data not shown). Sequence analysis of the 16S rRNA gene revealed the presence of nine distinct 16S rRNA alleles among the 30 isolates chosen for analysis, with differences being observed in 8 positions across the 1,435 bp examined. A tree built using the neighbor-joining method for the 16S rRNA gene (Fig. 5.1) showed a single branch that gave rise to all the *S. canis* alleles (SC1 to SC9). The *S. canis* type strain DSM 20715 16S rRNA sequence (GenBank accession number DQ303184) (41) is identical to the sequence of the *SC5* allele found in the isolates characterized in this work. The closest relative in the tree to the *S. canis* alleles was the GAS sequence used for comparison, which differed by 28 bp (98% identity) from the sequences of the *SC2* and *SC3* alleles. The single isolate identified as group L streptococcus by the API 20 Strep system had the *SC6* allele, together with another five isolates belonging to the same genetic lineage (see below). The GenBank accession numbers of the 16S rRNA gene sequences determined in the current study are given in Table S5.1 in the supplemental material.

Antimicrobial susceptibility

All isolates were susceptible to penicillin, cefotaxime, chloramphenicol, levofloxacin, vancomycin, quinupristin-dalfopristin, linezolid, and gentamicin. Tetracycline resistance was expressed by 23 isolates (27%) and 5 showed intermediate resistance, as determined by disk diffusion. Two tetracycline-resistant isolates were also resistant to macrolides and streptomycin (but not to gentamicin). Both isolates presented a macrolide, lincosamide, and streptogramin B (MLS_B) constitutive resistance

phenotype with MICs to erythromycin and clindamycin of >256 μg/ml, and they had MICs to streptomycin and gentamicin of >1,024 μg/ml and 3 μg/ml, respectively.

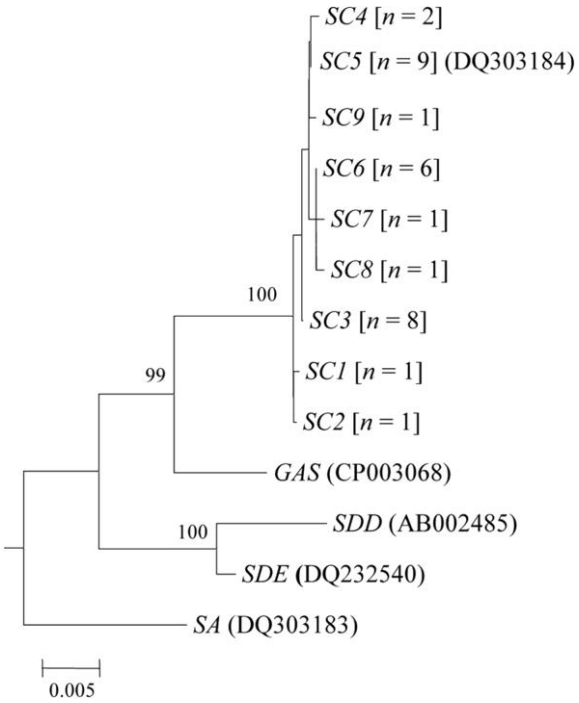

FIGURE 5.1. Neighbor-joining tree of the 16S rRNA

The *Streptococcus suis* type strain ATCC 43765 16S rRNA sequence (GenBank accession number DQ303193) was used to root the tree. Branches clustering sequences with greater that 75% bootstrap support (1,000 replicates) are indicated. The tree is drawn to scale, with branch lengths corresponding to the number of base substitutions per site. *S. canis* alleles determined in this work are named *SC1* to *SC9*. GenBank entries whose sequences were used to construct the tree are indicated in parentheses. The number of *S. canis* isolates presenting a given allele is indicated in brackets. *SC*, *S. canis*; *SDD*, *S. dysgalactiae* subsp. *dysgalactiae*; *SDE*, *S. dysgalactiae* subsp. *equisimilis*; *SA*, *S. agalactiae*; *GAS*, *S. pyogenes*.

The *tet*(M) and *tet*(O) genes were the sole determinants carried by 11 and 8 tetracycline-resistant isolates, respectively. One isolate simultaneously carried the *tet*(M) and *tet*(L) genes, and another one was positive for the *tet*(L) and *tet*(S) genes. Two resistant isolates were negative for all genes tested in this work, as were all the five isolates with intermediate resistance. The two isolates that were simultaneously resistant to tetracycline, macrolides, and streptomycin were *tet*(O) positive and carried the *erm*(B) and *aadA* genes. These isolates were recovered in 2004 and 2005 from dogs in Portugal, and they were clonally related, belonging to ST9 and being found in PFGE cluster B20 (see below). The remaining tetracycline-resistant isolates were found to be associated with various STs and PFGE clusters, including the main genetic lineages found among susceptible isolates.

***S. canis* clonal relationships inferred by MLST**

The novel MLST scheme proposed here enabled us to successfully characterize all *S. canis* isolates. Twenty-four STs (SID ± 95% CI, 0.887 ± 0.044) were found among the 86 isolates characterized in our study, from which 22 were novel STs (ST1 and ST5 to ST25). The number of alleles ranged from 5 (for *murI*, *mutS*, and *yqiZ*) to 9 (for *gki* and *xpt*). All alleles previously described in isolates recovered from human infections in the United States (26) were also present in isolates from house pets characterized in the current study, except for the *recP3* allele (GenBank accession number FJ238476), which was found solely in ST4. The numbering of the *S. canis* alleles found at the different *loci* and respective GenBank accession numbers are given in Table S5.1 in the supplemental material. The MLST data used in the current work are given in Table S5.2 in the supplemental material.

The goeBURST analysis of all 90 isolates that have been characterized by MLST to date (including the three isolates characterized previously and the *S. canis* FSL Z3-227 isolate for which MLST data were obtained from the genome sequence project) revealed that 17 STs were grouped into 5 CCs defined at the SLV level, while 8 STs were singletons (i.e., had no SLVs in the data set) (SID ± 95% CI, 0.804 ± 0.053) (Fig. 5.2). At the DLV level, 5 CCs and 3 singletons could be identified (SID ± 95% CI, 0.752 ± 0.054). All STs had at least one triple-locus-variant (TLV) in the data set, except for ST25, which shared no more than one allele with any other ST. The isolate having ST25 had unique *gki*, *gtr*, *mutS*, *recP*, and *yqiZ* alleles. This isolate also had a unique 16S rRNA sequence (allele *SC9* in Fig. 5.1) and was recovered from a cat in Germany. ST9 (n = 24), ST13 (n = 13), and ST1 (n = 11) were the most frequent STs in isolates from both Portugal and Germany and together accounted for more than half of the isolates. These STs were identified as the putative founders of the CCs in which they were included (Fig. 5.2). ST9 included the *S. canis* type strain DSM 20715 and the FSL Z3-227 isolate (both recovered from bovine mastitis), as well as five of the seven human isolates recovered in Portugal, and was the most frequent ST recovered from house pets in both geographic locations, including six of the eight cat isolates analyzed (Table 5.2). The single horse isolate analyzed presented ST10, a SLV of ST9. ST9 is also a SLV from ST3 detected in invasive infections in United States (26). Two additional human isolates characterized in this study represented ST12 and ST14. These STs did not share any allele with ST9 and were each other's SLVs, as well as the SLV of ST4, exclusively

recovered from a human infection in the North American study (26). With the exception of ST4, all S. canis STs found in human isolates were also found among isolates from house pets (Fig. 5.2A).

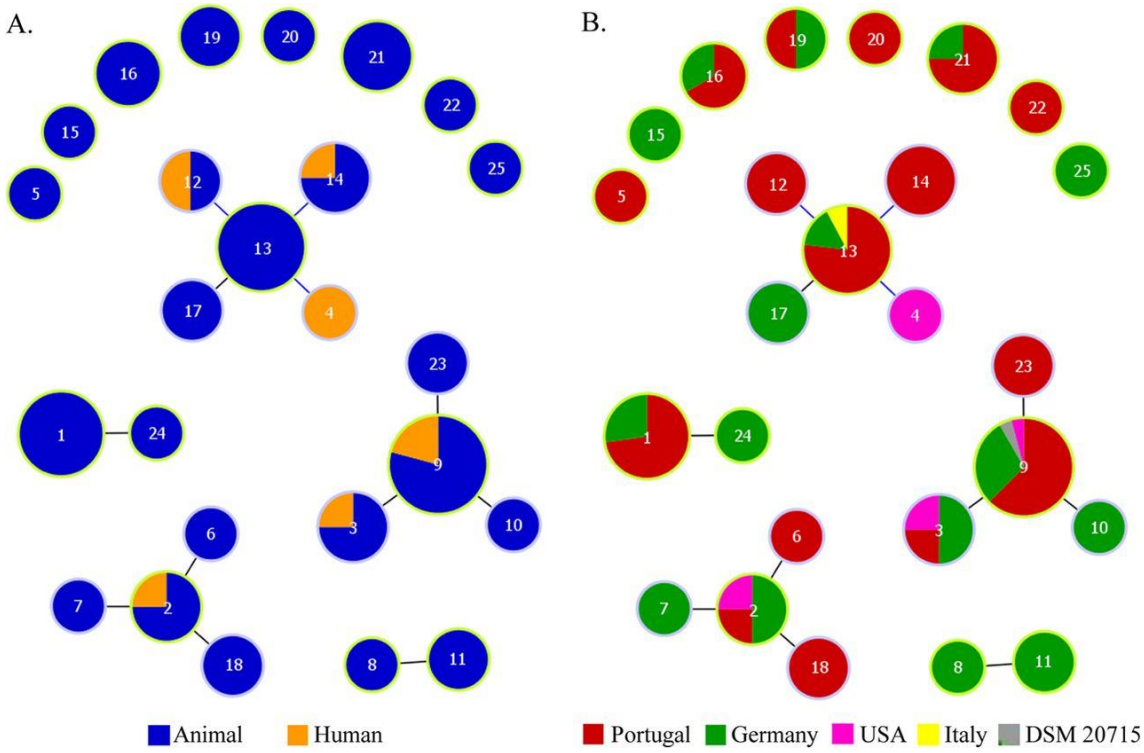

FIGURE 5.2. goeBURST diagram of the relationships between 90 *S. canis* isolates recovered from animal and human infections in distinct geographic locations

The size of each circle is proportional to the number of isolates with that particular ST on a logarithmic scale. STs assigned to the same CC at the SLV level are linked by straight lines. Whenever isolates of the same ST were recovered from different hosts (A) or from more than one location (B), the number of isolates with the same characteristic is proportional to the respective color. Putative CC founders are identified by a light green line.

ST9, ST2, and ST3 found in human isolates were also detected in house pets in both Portugal and Germany (Fig. 5.2B). The CCs formed included isolates from both Portugal and Germany, and only CC11, which included the smaller number of isolates (n = 3), was an exception. Three of the singletons were also found in the two countries. The isolates previously reported from the United States, as well as the single isolate recovered from Italy and the *S. canis* type strain, were distributed into the main CCs. No geographic segregation was evident from the goeBURST analysis.

TABLE 5.2. Distribution of PFGE clusters, *emm* types, and isolate origin among MLST STs and CCs

ST (no. of isolates)	CC[a]		PFGE cluster(s) (no. of isolates)[b]	*emm* type(s) (no. of isolates)[c]	Origin(s) (no. of isolates)
	SLV	DLV			
9 (23)	9	9	B_{20} (17), C_{10} (3), D_5 (3)	NT (23)	Dog (11), cat (6), human (5), cow[d] (1)
10 (1)	9	9	B_{20} (1)	NT (1)	Horse (1)
3 (3)	9	9	C_{10} (3)	NT (3)	Dog (3)
23 (2)	9	9	C_{10} (2)	NT (2)	Dog (2)
16 (3)	S	9	C_{10} (2), B_{20} (1)	NT (3)	Dog (2), cat (1)
22 (1)	S	9	G_2 (1)	NT (1)	Dog (1)
13 (13)	13	13	H_{25} (13)	NT (13)	Dog (13)
14 (4)	13	13	H_{25} (4)	NT (4)	Dog (3), human (1)
12 (2)	13	13	H_{25} (2)	NT (2)	Dog (1), human (1)
17 (2)	13	13	H_{25} (2)	NT (2)	Dog (2)
11 (2)	11	13	H_{25} (2)	NT (2)	Dog (2)
8 (1)	11	13	H_{25} (1)	NT (1)	Dog (1)
2 (3)	2	2	J_6 (2), I_2 (1)	NT (3)	Dog (3)
6 (1)	2	2	J_6 (1)	NT (1)	Dog (1)
7 (1)	2	2	J_6 (1)	NT (1)	Dog (1)
18 (2)	2	2	J_6 (2)	NT (2)	Dog (2)
1 (11)	1	1	A_5 (5), E_2 (2), F_2 (2), B_{20} (1), up (1)	*stG1389* (11)	Dog (11)
24 (1)	1	1	up (1)	*stG1389* (1)	Dog (1)
21 (4)	S	21	L_3 (3), G_2 (1)	NT (3), *stG1451* (1)	Dog (4)
15 (1)	S	21	H_{25} (1)	NT (1)	Dog (1)
20 (1)	S	21	I_2 (1)	*stG1451* (1)	Dog (1)
19 (2)	S	S	D_5 (2)	NT (2)	Dog (2)
5 (1)	S	S	K_2 (1)	*stG663* (1)	Dog (1)
25 (1)	S	S	K_2 (1)	NT (1)	Cat (1)

[a] CCs are identified by the identified putative founder. S, singleton.
[b] PFGE clusters are designated as indicated in the legend to Fig. 5.5A. up, unique profile.
[c] NT, nontypeable.
[d] *S. canis* type strain DSM 20715.

Phylogenetic analysis of *S. canis*

A tree was built by the neighbor-joining method for six concatenated *loci* and for the individual *loci* with the common data available for *S. canis*, SDE, and GAS. The concatenated sequences from each of the three species were clearly separated in branches with high bootstrap support (Fig. 5.3). In contrast to the 16S rRNA analysis (Fig. 5.1), this analysis identifies *S. canis* as a more divergent clade and SDE and GAS as more closely related. The majority of *S. canis* sequences were grouped together in a

branch with 99% support from which two branches diverged, one containing STs 20 and 21 and another with the more divergent ST25 (noted before for having unique alleles at five *loci*).

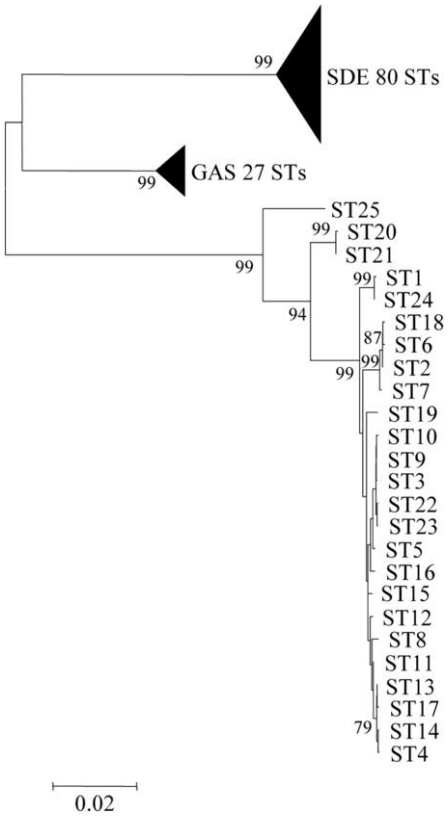

FIGURE 5.3. Neighbor-joining unrooted tree of concatenated sequences of the six common genes used in *S. canis*, SDE, and GAS MLST schemes

The tree was built with the methods indicated in Materials and Methods. *S. canis* STs are indicated ST1 to ST25. The concatenated STs from the two comparator species were obtained from the respective databases (http://www.mlst.net/) and consisted of 80 STs for *S. dysgalactiae* subsp. *equisimilis* and 27 STs for GAS (the group founders of CCs with at least five STs). A total of 2,700 bp for each sequence type was used in constructing the tree.

Trees were also built for the individual *loci*. A single branch with high bootstrap support (≥99%) containing only *S. canis* alleles was observed in the trees for all *loci*, independent of the overall tree topology, which was variable and showed *S. canis* alleles to be either more closely related to the alleles of GAS than to those of SDE (*gki*, *gtr*, and *murI*) or an outgroup of these two species (*mutS*, *recP*, and *xpt*). The *gki* and *recP* trees are shown as an example in Fig. 5.4 (see Fig. S5.1 in the supplemental material for the remaining trees). No *S. canis* alleles were shared with SDE or GAS, but two alleles, *gki7* and *recP7*, were divergent from all other *S. canis* alleles and were more similar to SDE alleles. The *gki7* allele had a 16-bp difference of the SDE *gki10* allele (97% identity), while sharing no more than 91% identity to any of the other *S. canis* alleles

(Fig. 5.4A). This allele was found in isolates of ST20 (*n* = 1) and ST21 (*n* = 4) noted above for their presence in a divergent branch. The *recP7* allele was also divergent from *S. canis* alleles, being almost identical to SDE *recP* alleles 12 (2-bp difference) and 13 (3-bp difference) across the 459-bp fragment of the *recP* gene used for analysis (Fig. 5.4B). *S. canis recP7* was exclusively found in the isolate of ST25.

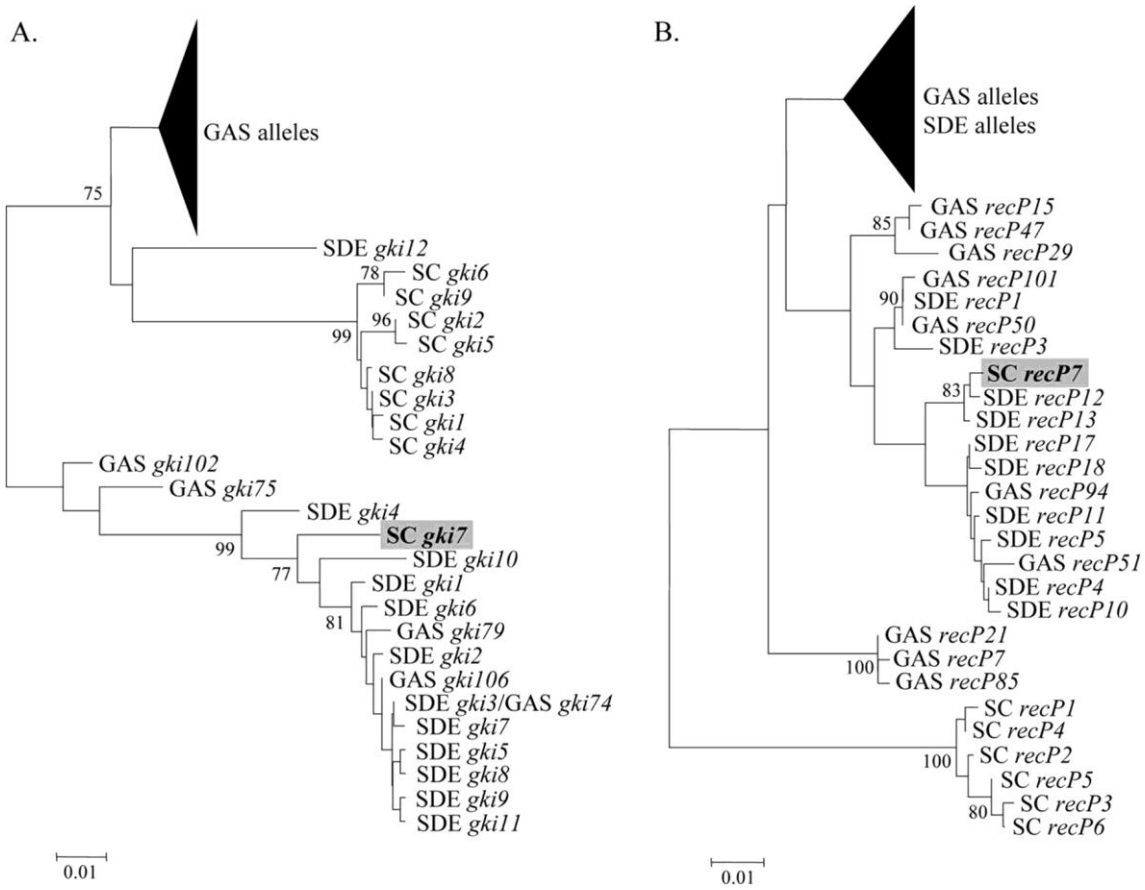

FIGURE 5.4. Neighbor-joining unrooted trees of *gki* (A) and *recP* alleles (B) of *S. canis*, SDE and GAS

The trees were built with the methods indicated in Materials and Methods. *gki* alleles are indicated *SC gki1* to *SC gki9*, and *recP* alleles are indicated *SC recP1* to *SC recP7*. The *gki* and *recP* sequences from the two comparator species were obtained from the respective databases (http://www.mlst.net/). The divergent *S. canis gki7* and *recP7* alleles are highlighted in the respective trees.

Diversity among *S. canis* isolates

Analysis of the PFGE patterns generated after digestion with SmaI revealed the presence of 12 clusters and two isolates with unique profiles (SID ± 95% CI, 0.842 ± 0.047) (Fig. 5.5A). The two larger PFGE clusters identified, H_{25} and B_{20}, correlated with the two most frequent STs, as they grouped mainly ST13 (*n* = 13) and ST9 (*n* = 17)

isolates. Table 5.2 shows the distribution of STs in the various PFGE clusters. Most PFGE clusters included more than one ST, and in some cases, the same ST was found among diverse PFGE clusters, resulting in an overall low level of association between the two typing techniques, as given by an AW coefficient between PFGE clusters and STs ($AW_{PFGE \rightarrow ST}$) of 0.364 (95% CI, 0.224 to 0.505) and an $AW_{ST \rightarrow PFGE}$ of 0.537 (95% CI, 0.368 to 0.706). The overall correspondence between PFGE cluster and CCs was significantly higher when CCs were formed at the DLV level [$AW_{PFGE \rightarrow CC(DLV)}$ = 0.880 (95% CI, 0.754 to 1)], reflecting the fact that that a significant fraction of STs included

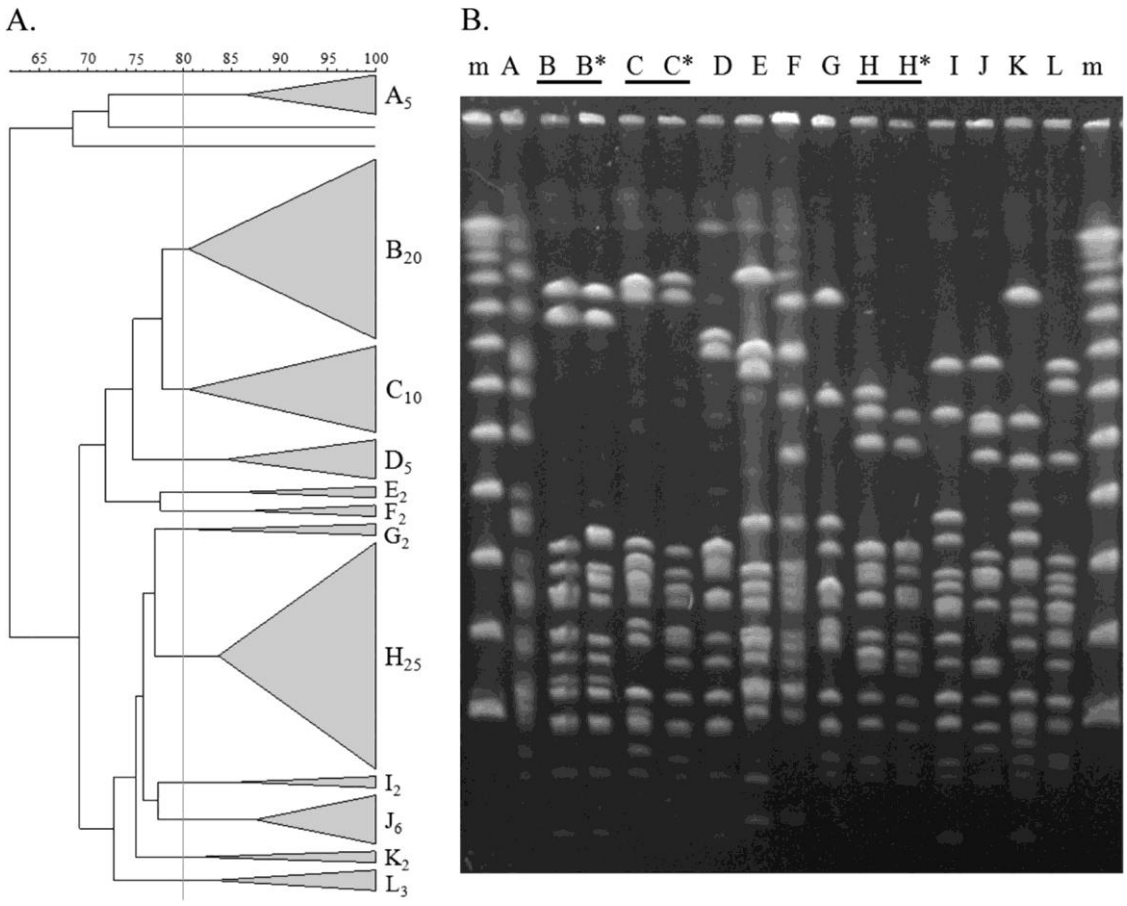

FIGURE 5.5. PFGE SmaI macrorestriction profile analysis of *S. canis* isolates from animal and human infections

(A) Dendrogram showing UPGMA cluster analysis of the PFGE profiles of the 86 *S. canis* isolates studied. Dice coefficients (percentages) are indicated on the scale above the dendrogram. Each PFGE cluster (defined as a group of ≥2 isolates with a Dice coefficient of ≥80%) is represented by a triangle proportional to the number of isolates included in the cluster. PFGE clusters are designated by capital letters and a subscript number indicating the number of isolates included in the cluster.

(B) PFGE profiles of representatives of *S. canis* clusters defined. Capital letters above the lanes correspond to cluster designations; pairs of isolates recovered from humans (indicated by an asterisk) and house pets belonging to the same PFGE cluster and with identical STs are underlined and include two ST9/B_{20} isolates, two ST9/C_{10} isolates, and two ST13/H_{25} isolates; lane m, bacteriophage lambda ladder PFGE marker (New England BioLabs, Beverly, MA).

in the same PFGE cluster belonged to the same CC (Table 5.2). An example is the inclusion of STs 13, 8, 11, 12, 14, and 17, belonging to CC13 defined at the DLV level, in the larger PFGE clone, H_{25} (Table 5.2).

Isolates recovered from humans and house pets were grouped together in the major PFGE clusters, similarly to what was noted above for MSLT-defined groups. Figure 5.5B shows three examples of paired PFGE profiles of human and house pet isolates belonging to the same PFGE cluster.

An *emm* gene could be amplified from 15 isolates (17%). None of the human isolates yielded a product when tested for the presence of the *emm*-like gene. All the *emm*-like positive isolates were recovered from dogs. Twelve isolates were *stG1389*, two had a novel *emm* type (*stG1451*), and one was *stG663*. The *stG1389* type was found in all isolates belonging to STs 1 (n = 11) and 24 (n = 1), both belonging to CC1. The novel *emm* type, *stG1451*, was found in one ST20 isolate and one ST21 isolate, both singletons at the SLV level that were double-locus-variants of each other. The other three ST21 isolates did not amplify the *emm*-like gene. *stG663* was found in an ST5 isolate (also a singleton). *emm*-positive isolates were found scattered in the PFGE dendrogram, including *stG1389* isolates which, despite belonging to a single MLST CC, had mostly unrelated PFGE profiles (Table 5.2).

DISCUSSION

S. canis, a widely distributed house pet colonizer and pathogen, is increasingly being reported in human infection (20, 21, 23). In industrialized countries, a trend toward closer contact and increased cohabitation in households with traditional pets such as dogs and cats has been noted (42, 43), making transmission of microorganisms between house pets and humans more likely to occur. These observations show that the zoonotic potential of *S. canis* should not be dismissed and prompted us to explore if *S. canis* isolates found in house pets and humans form a single population.

Many previous epidemiologic studies of human infection by GCGS failed to detect *S. canis* (28, 44, 45). In Portugal, an ongoing survey of human infections by large-colony-forming GCGS did not detect any *S. canis* isolates prior to 2005, while over 300 SDE isolates were recovered in the same period (46). The seven *S. canis* isolates recovered from both invasive and non-invasive human infections in Portugal between 2006 and 2010 described herein represent approximately 1.2% of all large-colony-forming GCGS isolates recovered during this time period and 3.2% of the large-colony-forming GCGS invasive isolates (unpublished data). This is a prevalence similar to the one reported for invasive isolates in the United States (19). In sharp contrast to our findings, a retrospective study carried out in a French hospital was unique, in that the number of infections caused by *S. canis* exceeded the number of infections caused by SDE and *S. canis* accounted for 1% of all isolated streptococci (20). That report also identified an unusually high number of *S. canis* isolates from nonsterile sites, with many being recovered together with other bacteria and some being recovered from patients with no signs of infection (20). In our study, three of the seven human isolates were recovered from nonsterile sites, including two from the respiratory tract (Table 5.1). The patients did present signs of respiratory infection, but we cannot definitively establish a pathogenic role for these isolates. These observations raise the possibility of *S. canis* colonization in humans, but such a role remains controversial.

All the group G isolates described in the current work were identified as *S. canis* by combining phenotypic and genotypic tests. The analysis of the 23S rRNA and 16S rRNA showed that we were able to correctly identify more than 99% of the isolates by using the API 20 Strep system alone. The high number of distinct API profiles obtained is in agreement with previous observations of considerable biochemical variation

among *S. canis* isolates (21), but all profiles were identical to the *S. canis* profiles obtained by the same method and reported elsewhere (18, 21). Taken together, our results support the usefulness of the API 20 Strep system for *S. canis* identification.

The phylogenetic tree built for the 16S rRNA gene showed that all *S. canis* alleles formed a single branch well resolved from the other beta-hemolytic streptococcal species included. *S. canis* alleles were more closely related to those of GAS, concordant with the most frequent tree topology observed in 16S rRNA analysis of streptococci (1). The genotypic analysis also showed that the group L streptococcus identification obtained for a single isolate by the API 20 Strep system was erroneous. Among *Streptococcus* species, the group L carbohydrate is found only among isolates in the *S. dysgalactiae* taxon, and all *S. canis* isolates identified to date present Lancefield group G antigen (1), as did this isolate. In addition, this isolate yielded a product specific for *S. canis* in the 23S rRNA PCR assay and had a 16S rRNA allele characteristic of *S. canis*. The results of molecular typing also showed that it was part of a well-defined *S. canis* genetic lineage, namely, the ST1, *emm* type *stG1389* present in PFGE cluster A5 (Table 5.2). Taken together, these results unambiguously identify this isolate as *S. canis*.

Like other beta-hemolytic species, *S. canis* isolates are usually susceptible to antimicrobial agents, but occasional resistance to tetracycline, fluoroquinolones, macrolides and lincosamides, gentamicin, and chloramphenicol has been reported (6, 21, 23, 47). The rate of tetracycline resistance (27%) for house pet isolates found in our study is similar to the one reported from dog isolates in Denmark (48) and cat isolates in Belgium (49). Resistance to tetracycline was significantly disseminated in *S. canis*, as confirmed by the polyclonal origin of the isolates expressing it. On the contrary, resistance to erythromycin, clindamycin, and streptomycin was minimal, being restricted to two genetically related isolates. Erythromycin and clindamycin resistance rates above 10% have been reported for *S. canis* (6, 48, 49) and for group G isolates recovered from human infections (50, 51). Also in contrast to our observations, resistance to gentamicin but not to streptomycin has frequently been detected in *S. canis* (6, 21, 23, 48). We could not detect resistance to levofloxacin in *S. canis*, in sharp contrast to our own observations on SDE isolates recovered from human infections in Portugal, for which a resistance rate of 12% was found (46).

Information on the genetic determinants of resistance carried by *S. canis* is lacking in the literature. Tetracycline resistance in our isolates was conferred mainly by *tet*(M) and

tet(O), which are also the genes found the most frequently in other beta-hemolytic streptococci of human and animal origin (32, 52). The presence of the *tet*(L) and *tet*(S) genes has also been noted, but at lower frequencies (32, 53, 54). The detection of two tetracycline-resistant isolates negative for the presence of the genes tested in this work may indicate that other determinants could be carried by *S. canis* isolates, which is not surprising, given the known diversity of *tet* genes (55). The *erm*(B) gene conferring constitutive MLS_B resistance in our isolates has also been detected in SDE (51). Our isolates were resistant to streptomycin while maintaining gentamicin susceptibility; accordingly, these isolates were shown to carry the *aadA* gene, previously detected in Gram-negative bacteria and *Enterococcus faecalis* (33).

Several typing techniques have been employed in the past to characterize *S. canis*, but most studies relied on a limited number of isolates. We propose a new *S. canis* MLST scheme and demonstrated its effectiveness by typing a larger collection of isolates. The goeBURST analysis showed that most STs were grouped into 5 CCs, accounting for more than 80% of the isolates and including the unrelated STs described in the North American study (26). Moreover, the putative founders of the main CCs defined by MLST were identified in Portugal and Germany, where most *S. canis* isolates were recovered, and some of the singletons also included isolates recovered in both countries. Taken together, the goeBURST analysis showed that the observed structure of the *S. canis* population is not limited to a single geographic location and that *S. canis* clones are geographically disseminated, as previously observed for human SDE (34).

The sequence data generated by MLST were also used to make a comparative phylogenetic analysis of *S. canis* with SDE and GAS. The tree topologies that we found for the different *loci* are in accordance with what has been reported (26), with *S. canis* or SDE alleles being more closely related to GAS, depending on the similarity observed between the last two species, which changes the tree topology (26, 34). Although a previous study based on multilocus sequence analysis of these three species showed a close genetic relationship and likely recent descent from a common ancestor (56), enough sequence divergence has accumulated so that the *S. canis* alleles and STs form discrete branches that are well separated from those of the other two species (26, 56). In view of this, the detection of one *gki* allele and one *recP* allele of *S. canis* more similar to SDE alleles than to any other *S. canis* allele can be most parsimoniously explained by

their acquisition through recombination. The occurrence of DNA exchange involving SDE and GAS housekeeping genes is documented (26, 34, 57). Genomic analysis has recently shown significant lateral gene transfer from GAS to *Streptococcus equi* subsp. *equ*i (35), while the data presented here support the hypothesis that SDE may be the major donor in the events involving *S. canis*. The detection in the present study of these events was probably made possible due to the increase in the number of isolates analyzed. Future studies which enlarge the known diversity of MLST alleles and STs of *S. canis* will provide further insights into the frequency of these events, allowing us to better evaluate their potential importance in the evolution of these species.

The results obtained by MLST were supported by characterizing the same isolates by PFGE. Although previous studies that characterized *S. canis* isolates from house pets also using PFGE SmaI profiles found a diverse population (13, 27), the presence of well-defined clones by both PFGE and MLST in the current study points to the existence of stable genetic lineages in the *S. canis* population which can be detected by both typing techniques. The overall agreement between PFGE and MLST was weak, as denoted by the low values obtained for the adjusted Wallace coefficient between PFGE clusters and STs or CCs, and only the analysis at the DLV level significantly maximized the association of genetically related STs with PFGE clusters (Table 5.2). This observation is not unique to *S. canis*, as poor concordance between typing methods has also been noted in the past for SDE, namely, in studies that compared *emm* typing results with either PFGE (28) or MLST (26, 34) typing results. The reasons for these observations are not clear, but unlike MLST, PFGE targets the entire genome and measures variation that could be accumulating more rapidly (58).

Although a zoonotic origin of *S. canis* has been assumed since the first reports of human infection (18, 21), most papers have reported the independent and exclusive characterization of either human or animal isolates. The MLST characterization performed in the current study showed that all the *S. canis* isolates from human infections (from the United States and Portugal) characterized by this technique to date have identical alleles and belong to the most prevalent STs/CCs present among house pet isolates, a strong indication that *S. canis* isolates are not segregated by host. Furthermore, the inclusion of epidemiologically unrelated human and house pet isolates in common PFGE clusters further supported the MLST findings. Taken together, these observations lead us to conclude that *S. canis* from house pets and humans form a single

population. The two *S. canis* isolates from bovine mastitis and the horse isolate had STs identical or related to isolates of house pets and humans, indicating that this observation may be extended to *S. canis* isolates found in other animal species. As both MLST and PFGE analysis showed that human isolates do not represent separate lineages or unusual clones, one can infer that the ability to infect the human host is not restricted to a few *S. canis* strains but, rather, that several genetic lineages have that pathogenic potential.

The major role played by the M protein in GAS pathogenesis and the virtually ubiquitous presence of *emm*-like genes among SDE isolates infecting humans raised the possibility that at least the *S. canis* isolates causing infections in humans would carry *emm*-like genes. However, only a small proportion of *S. canis* isolates were successfully *emm* typed, and in none of the human isolates was this gene successfully amplified. Despite the successful application of the currently used *emm* typing protocol for both GAS and SDE, many previous studies failed to amplify an *emm*-like gene from most *S. canis* isolates (19, 23), in agreement with our observations. Information accessible at the CDC *emm* typing database shows that both the *stG1389* and *stG663 emm* types were previously identified in *S. canis* (http://www.cdc.gov/ncidod/biotech/strep/types_emm103-124.htm), while *stG1451* is a novel *emm* type. None of these *emm* types have been observed among SDE isolates from Portugal (46) or elsewhere.

The data from the current study provide new insights into the clonal relationships between *S. canis* isolates infecting distinct hosts. Our report shows for the first time that several prominent *S. canis* lineages found among house pets also have the potential to cause invasive infections in humans. By demonstrating that the same MLST- and PFGE-based genetic lineages infect both house pets and humans, we provide decisive evidence for the zoonotic origin of *S. canis*.

ACKNOWLEDGMENTS

We thank Christel Simon, Andrea Schmidt, and Marlene Delgado for their excellent technical assistance and João A. Carriço for help with genomic analysis of strain FSL Z3-227.

We thank the members of the Portuguese Group for the Study of Streptococcal Infections for the surveillance of *S. canis* human infections: Teresa Vaz, Marília Gião, Rui Ferreira, and Iryna Klyeshtorna (Centro Hospitalar do Barlavento Algarvio), Ana Buschy Fonseca (Hospital de Cascais), Henrique Oliveira (Centro Hospitalar de Coimbra), Ana Cristina Silva, Hermínia Costa, Maria Fátima Silva, and Maria Amélia Afonso (Centro Hospitalar de Entre Douro e Vouga), Margarida Pinto, Odete Chantre, João Marques, Isabel Peres, Isabel Daniel, and Cristina Marcelo (Centro Hospitalar de Lisboa Central), Lurdes Monteiro and Luís Marques Lito (Centro Hospitalar Lisboa Norte), Teresa Marques, Maria Ana Pessanha, and Elsa Gonçalves (Centro Hospitalar Lisboa Ocidental), Paulo Lopes, Luísa Felício, and Angelina Lameirão (Centro Hospitalar de Vila Nova de Gaia/Espinho), Ana Paula Mota Vieira and Margarida Tomaz (Centro Hospitalar do Alto Ave), Rosa Bento (Centro Hospitalar do Baixo Alentejo), Maria Helena Ramos and Ana Apula Castro (Centro Hospitalar do Porto), Fernando Fonseca (Centro Hospitalar da Póvoa do Varzim/Vila do Conde), Ana Paula Castro (Centro Hospitalar Trás-os-Montes e Alto Douro), Graça Ribeiro, Luísa Boaventura, Catarina Chaves, and Teresa Reis (Hospitais da Universidade de Coimbra), Nuno Canhoto and Teresa Afonso (Hospital Central do Funchal), Teresa Pina and Helena Peres (Hospital Curry Cabral, Lisbon), Ilse Fontes and Paulo Martinho (Hospital de Santa Luzia, Elvas), Ana Domingos and Gina Marrão (Hospital de Santo André, Leiria), Manuela Ribeiro and Helena Gonçalves (Hospital de São João, Porto), Maria Alberta Faustino, Maria Carmen Iglesias, and Adelaide Alves (Hospital de Braga), Maria Paula Pinheiro and R. Semedo (Hospital José Maria Grande, Portalegre), Adriana Coutinho (Hospital do Espírito Santo, Évora), Luísa Cabral and Olga Neto (Hospital dos SAMS, Lisbon), Luísa Sancho (Hospital Fernando da Fonseca, Amadora/Sintra), José Diogo, Ana Rodrigues, and Isabel Nascimento (Hospital Garcia de Orta, Almada), Elmano Ramalheira and Raquel Diaz (Hospital Infante D. Pedro, Aveiro), José Miguel Ribeiro, Isabel Vale, and Ana Carvalho (Hospital de São Teotónio, Viseu), Maria Antónia Read, Margarida Monteiro, and Valquíria Alves (Hospital Pedro Hispano, Matosinhos), Engrácia Raposo, Maria Lurdes Magalhães, Helena Rochas, and Anabela

Silva (Instituto Nacional de Saúde Ricardo Jorge, Porto), Margarida Rodrigues (Hospital Reynaldo dos Santos, Vila Franca de Xira), Eulália Carvalho and Karine Hyde (Hospital do Divino Espírito Santo, Ponta Delgada), and Clotilde Roldão (Hospital Distrital de Abrantes).

This work was partially supported by the Fundação para a Ciência e Tecnologia, Portugal (PTDC/CVT-EPI/4695/2012) and the International Research Training Group (IRTG) 1673 from the German Research Foundation (DFG).

J. Melo-Cristino has received research grants administered through his university and received honoraria for consulting and serving on the speakers' bureaus of Pfizer, Bial, GlaxoSmithKline, and Novartis. M. Ramirez has received honoraria for consulting and serving on the speakers' bureau of Pfizer. The other authors declare no conflict of interest. No company or financing body had any interference in the decision to publish.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: M. D. Pinho, C. Pomba, J. Melo-Cristino and M. Ramirez. Performed the experiments: M. D. Pinho and S. C. Matos. Analyzed the data: M. D. Pinho, J. Melo-Cristino and M. Ramirez. Contributed reagents/materials/analysis tools: C. Pomba, A. Lübke-Becker, L. H. Wieler, S. Preziuso, J. Melo-Cristino and M. Ramirez. Wrote the paper: M. D. Pinho, C. Pomba, A. Lübke-Becker, L. H. Wieler, S. Preziuso, J. Melo-Cristino and M. Ramirez.

REFERENCES

1. **Facklam R.** 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin. Microbiol. Rev. **15**:613–630.
2. **Devriese LA, Hommez J, Kilpper-Balz R, Schleifer KH.** 1986. *Streptococcus canis* sp. *nov*.: a species of group G streptococci from animals. Int. J. Syst. Bacteriol. **36**:422–425.
3. **Clark RB, Berrafati JF, Janda JM, Bottone EJ.** 1984. Biotyping and exoenzyme profiling as an aid in the differentiation of human from bovine group G streptococci. J. Clin. Microbiol. **20**:706–710.
4. **Clemetson LL, Ward AC.** 1990. Bacterial flora of the vagina and uterus of healthy cats. J. Am. Vet. Med. Assoc. **196**:902–906.
5. **Devriese LA, Cruz Colque JI, De Herdt P, Haesebrouck F**. 1992. Identification and composition of the tonsillar and anal enterococcal and streptococcal flora of dogs and cats. J. Appl. Bacteriol. **73**:421–425.
6. **Lyskova P, Vydrzalova M, Mazurova J.** 2007. Identification and antimicrobial susceptibility of bacteria and yeasts isolated from healthy dogs and dogs with otitis externa. J. Vet. Med. A Physiol. Pathol. Clin. Med. **54**:559–563.
7. **Watts JR, Wright PJ, Whithear KC.** 1996. Uterine, cervical and vaginal microflora of the normal bitch throughout the reproductive cycle. J. Small Anim. Pract. **37**:54–60.
8. **Hariharan H, Matthew V, Fountain J, Snell A, Doherty D, King B, Shemer E, Oliveira S, Sharma RN**. 2011. Aerobic bacteria from mucous membranes, ear canals, and skin wounds of feral cats in Grenada, and the antimicrobial drug susceptibility of major isolates. Comp. Immunol. Microbiol. Infect. Dis. **34**:129–134.
9. **Jepson PD, Baker JR, Kuiken T, Simpson VR, Kennedy S, Bennett PM**. 2000. Pulmonary pathology of harbour porpoises (*Phocoena phocoena*) stranded in England and Wales between 1990 and 1996. Vet. Rec. **146**:721–728.
10. **Murase T, Morita T, Sunagawa Y, Sawada M, Shimada A, Sato K, Hikasa Y**. 2003. Isolation of *Streptococcus canis* from a Japanese raccoon dog with fibrinous pleuropneumonia. Vet. Rec. **153**:471–472.
11. **Lamm CG, Ferguson AC, Lehenbauer TW, Love BC**. 2010. Streptococcal infection in dogs: a retrospective study of 393 cases. Vet. Pathol. **47**:387–395.
12. **Sykes JE, Kittleson MD, Pesavento PA, Byrne BA, MacDonald KA, Chomel BB**. 2006. Evaluation of the relationship between causative organisms and clinical characteristics of infective endocarditis in dogs: 71 cases (1992-2005). J. Am. Vet. Med. Assoc. **228**:1723–1734.
13. **DeWinter LM, Prescott JF**. 1999. Relatedness of *Streptococcus canis* from canine streptococcal toxic shock syndrome and necrotizing fasciitis. Can. J. Vet. Res. **63**:90–95.
14. **Pesavento PA, Bannasch MJ, Bachmann R, Byrne BA, Hurley KF**. 2007. Fatal *Streptococcus canis* infections in intensively housed shelter cats. Vet. Pathol. **44**:218–221.

15. **Tikofsky LL, Zadoks RN.** 2005. Cross-infection between cats and cows: origin and control of *Streptococcus canis* mastitis in a dairy herd. J. Dairy Sci. **88**:2707–2713.

16. **Chaffer M, Friedman S, Saran A, Younis A.** 2005. An outbreak of *Streptococcus canis* mastitis in a dairy herd in Israel. N. Z. Vet. J. **53**:261–264.

17. **Hassan AA, Akineden O, Usleber E.** 2005. Identification of *Streptococcus canis* isolated from milk of dairy cows with subclinical mastitis. J. Clin. Microbiol. **43**:1234–1238.

18. **Bert F, Lambert-Zechovsky N**. 1997. Septicemia caused by *Streptococcus canis* in a human. J. Clin. Microbiol. **35**:777–779.

19. **Broyles LN, Van Beneden C, Beall B, Facklam R, Shewmaker PL, Malpiedi P, Daily P, Reingold A, Farley MM**. 2009. Population-based study of invasive disease due to beta-hemolytic streptococci of groups other than A and B. Clin. Infect. Dis. **48**:706–712.

20. **Galpérine T, Cazorla C, Blanchard E, Boineau F, Ragnaud Neau J-MD**. 2007. *Streptococcus canis* infections in humans: retrospective study of 54 patients. **J. Infect. 55**:23–26.

21. **Lam MM, Clarridge JE, Young EJ, Mizuki S**. 2007. The other group G *Streptococcus*: increased detection of *Streptococcus canis* ulcer infections in dog owners. J. Clin. Microbiol. **45**:2327–2329.

22. **Takeda N, Kikuchi K, Asano R, Harada T, Totsuka K, Sumiyoshi T, Uchiyama T, Hosoda S**. 2001. Recurrent septicemia caused by *Streptococcus canis* after a dog bite. Scand. J. Infect. Dis. **33**:927–928.

23. **Whatmore AM, Engler KH, Gudmundsdottir G, Efstratiou A.** 2001. Identification of isolates of *Streptococcus canis* infecting humans. J. Clin. Microbiol. **39**:4196–4199.

24. **Efstratiou A, Colman G, Hahn G, Timoney JF, Boeufgras JM, Monget D**. 1994. Biochemical differences among human and animal streptococci of Lancefield group C or group G. J. Med. Microbiol. **41**:145–148.

25. **Kawata K, Anzai T, Senna K, Kikuchi N, Ezawa A, Takahashi T**. 2004. Simple and rapid PCR method for identification of streptococcal species relevant to animal infections based on 23S rDNA sequence. FEMS Microbiol. Lett. **237**:57–64.

26. **Ahmad Y, Gertz RE, Li Z, Sakota V, Broyles LN, Van Beneden C, Facklam R, Shewmaker PL, Reingold A, Farley MM, Beall BW**. 2009. Genetic relationships deduced from *emm* and multilocus sequence typing of invasive *Streptococcus dysgalactiae* subsp. *equisimilis* and *S. canis* recovered from isolates collected in the United States. J. Clin. Microbiol. **47**:2046–2054.

27. **Kruger EF, Byrne BA, Pesavento P, Hurley KF, Lindsay LL, Sykes JE.** 2010. Relationship between clinical manifestations and pulsed-field gel profiles of *Streptococcus canis* isolates from dogs and cats. Vet. Microbiol. **146**:167–171.

28. **Pinho MD, Melo-Cristino J, Ramirez M.** 2006. Clonal relationships between invasive and non-invasive Lancefield group C and G streptococci and *emm*-specific differences in invasiveness. J. Clin. Microbiol. **44:**841–846.

29. **Clinical and Laboratory Standards Institute**. 2012. Performance standards for antimicrobial susceptibility testing; 22nd informational supplement. CLSI document M100-S22. Clinical and Laboratory Standards Institute, Wayne, PA.

30. **Figueira-Coelho J, Ramirez M, Salgado MJ, Melo-Cristino J.** 2004. *Streptococcus agalactiae* in a large Portuguese teaching hospital: antimicrobial susceptibility, serotype distribution, and clonal analysis of macrolide-resistant isolates. Microb. Drug Resist. **10**:31–36.

31. **Trzcinski K, Cooper BS, Hryniewicz W, Dowson CG**. 2000. Expression of resistance to tetracyclines in strains of methicillin-resistant *Staphylococcus aureus*. J. Antimicrob. Chemother. **45**:763–770.

32. **Haenni M, Saras E, Bertin S, Leblond P, Madec Payot J-YS.** 2010. Diversity and mobility of integrative and conjugative elements in bovine isolates of *Streptococcus agalactiae*, *S. dysgalactiae* subsp. *dysgalactiae*, and *S. uberis*. Appl. Environ. Microbiol. **76**:7957–7965.

33. **Clark NC, Olsvik O, Swenson JM, Spiegel CA, Tenover FC**. 1999. Detection of a streptomycin/spectinomycin adenylyltransferase gene (*aadA*) in *Enterococcus faecalis*. Antimicrob. Agents Chemother. **43**:157–160.

34. **McMillan DJ, Bessen DE, Pinho M, Ford C, Hall GS, Melo-Cristino J, Ramirez M.** 2010. Population genetics of *Streptococcus dysgalactiae* subspecies *equisimilis* reveals widely dispersed clones and extensive recombination. PLoS One **5**:e11741.

35. **Lefébure T, Richards VP, Lang P, Pavinski-Bitar P, Stanhope MJ**. 2012. Gene repertoire evolution of *Streptococcus pyogenes* inferred from phylogenomic analysis with *Streptococcus canis* and *Streptococcus dysgalactiae*. PLoS One **7:**e37607.

36. **Francisco AP, Bugalho M, Ramirez M, Carriço JA**. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics **10**:152.

37. **Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA**. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics **13**:87.

38. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S**. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.

39. **Carriço JA, Silva-Costa C, Pinto FR, De Lencastre H, Almeida JS, Ramirez M.** 2006. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. J. Clin. Microbiol. **44**:2524–2532.

40. **Severiano A, Pinto FR, Ramirez M, Carriço JA**. 2011. Adjusted Wallace coefficient as a measure of congruence between typing methods. J. Clin. Microbiol. **49**:3997–4000.

40a. **Jolley KA, Maiden MCJ.** 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics **11**:595.

41. **Bekal S, Gaudreau C, Laurence RA, Simoneau E, Raynal L**. 2006. *Streptococcus pseudoporcinus* sp. *nov*., a novel species isolated from the genitourinary tract of women. J. Clin. Microbiol. **44**:2584–2586.

42. **Chomel BB, Sun B**. 2011. Zoonoses in the bedroom. Emerging Infect. Dis. **17**:167–172.

43. **Walther B, Hermes J, Cuny C, Wieler LH, Vincze S, Abou Elnaga Y, Stamm I, Kopp PA, Kohn B, Witte W, Jansen A, Conraths FJ, Semmler T, Eckmanns T, Lübke-Becker A.** 2012. Sharing more than friendship—nasal colonization with coagulase-positive staphylococci (CPS) and co-habitation aspects of dogs and their owners. PLoS One **7**:e35197.

44. **Efstratiou A.** 1997. Pyogenic streptococci of Lancefield groups C and G as pathogens in man. Soc. Appl. Bacteriol. Symp. Ser. **26 Suppl**:72–79.

45. **Rantala S, Vuopio-Varkila J, Vuento R, Huhtala H, Syrjänen J.** 2009. Clinical presentations and epidemiology of beta-haemolytic streptococcal bacteraemia: a population-based study. Clin. Microbiol. Infect. **15**:286–288.

46. **Pinho MD, Melo-Cristino J, Ramirez M**. 2010. Fluoroquinolone resistance in *Streptococcus dysgalactiae* subsp. *equisimilis* and evidence for a shared global gene pool with *Streptococcus pyogenes*. Antimicrob. Agents Chemother. **54**:1769–1777.

47. **Awji EG, Damte D, Lee S-J, Lee J-S, Kim Y-H, Park S-C.** 2012. The in vitro activity of 15 antimicrobial agents against bacterial isolates from dogs. J. Vet. Med. Sci. **74**:1091–1094.

48. **Pedersen K, Pedersen K, Jensen H, Finster K, Jensen VF, Heuer OE**. 2007. Occurrence of antimicrobial resistance in bacteria from diagnostic samples from dogs. J. Antimicrob. Chemother. **60**:775–781.

49. **Moyaert H, De Graef EM, Haesebrouck F, Decostere A.** 2006. Acquired antimicrobial resistance in the intestinal microbiota of diverse cat populations. Res. Vet. Sci. **81**:1–7.

50. **Merino Díaz L, Torres Sánchez MJ, Aznar Martín J**. 2008. Prevalence and mechanisms of erythromycin and clindamycin resistance in clinical isolates of beta-haemolytic streptococci of Lancefield groups A, B, C and G in Seville, Spain. Clin. Microbiol. Infect. **14**:85–87.

51. **Sunaoshi K, Murayama SY, Adachi K, Yagoshi M, Okuzumi K, Chiba N, Morozumi M, Ubukata K**. 2010. Molecular *emm* genotyping and antibiotic susceptibility of *Streptococcus dysgalactiae* subsp. *equisimilis* isolated from invasive and non-invasive infections. J. Med. Microbiol. **59**:82–88.

52. **Soedarmanto I, Schwarz S, Liebisch B, Lämmler C**. 1995. Tetracycline resistance determinants among streptococci of serological group G and L. Vet. Microbiol. **45**:331–337.

53. **Liu L-C, Tsai J-C, Hsueh P-R, Tseng S-P, Hung W-C, Chen H-J, Teng L-J**. 2008. Identification of *tet*(S) gene area in tetracycline-resistant *Streptococcus dysgalactiae* subsp. *equisimilis* clinical isolates. J. Antimicrob. Chemother. **61**:453–455.

54. **Rodriguez-Avial I, Rodriguez-Avial C, Culebras E, Picazo JJ**. 2003. Distribution of tetracycline resistance genes *tet*(M), *tet*(O), *tet*(L) and *tet*(K) in blood isolates of *viridans*

group streptococci harbouring *erm*(B) and *mef*(A) genes. Susceptibility to quinupristin/dalfopristin and linezolid. Int. J. Antimicrob. Agents **21**:536–541.

55. **Roberts MC**. 2005. Update on acquired tetracycline resistance genes. FEMS Microbiol. Lett. **245**:195–203.

56. **Jensen A, Kilian M**. 2012. Delineation of *Streptococcus dysgalactiae*, its subspecies, and its clinical and phylogenetic relationship to *Streptococcus pyogenes*. J. Clin. Microbiol. **50**:113–126.

57. **McMillan DJ, Kaul SY, Bramhachari PV, Smeesters PR, Vu T, Karmarkar MG, Shaila MS, Sriprakash KS**. 2011. Recombination drives genetic diversification of *Streptococcus dysgalactiae* subspecies *equisimilis* in a region of streptococcal endemicity. **PLoS One 6**:e21346.

58. **Enright MC, Spratt BG**. 1999. Multilocus sequence typing. Trends Microbiol. **7**:482–487.

SUPPLEMENTAL MATERIAL

TABLE S5.1. GenBank accession numbers of 16S rRNA gene, MLST and novel *emm* type alleles

Gene	Allele	GenBank acession number	Reference
16S rRNA	1	JX876612	This study
16S rRNA	2	JX876613	This study
16S rRNA	3	JX876614	This study
16S rRNA	4	JX876615	This study
16S rRNA	5	DQ303184	41
16S rRNA	6	JX876616	This study
16S rRNA	7	JX876617	This study
16S rRNA	8	JX876618	This study
16S rRNA	9	JX876619	This study
gki	1	EU768807	26
gki	2	EU768805	26
gki	3	EU768804	26
gki	4	FJ238467	26
gki	5	JX876620	This study
gki	6	JX876621	This study
gki	7	JX876622	This study
gki	8	JX876623	This study
gki	9	JX876624	This study
gtr	1	FJ238468	26
gtr	2	EU768802	26
gtr	3	EU999155	26
gtr	4	JX876625	This study
gtr	5	JX876626	This study
gtr	6	JX876627	This study
gtr	7	JX876628	This study
murI	1	EU999157	26
murI	2	FJ151278	26
murI	3	FJ238469	26
murI	4	FJ151277	26
murI	5	JX876629	This study
mutS	1	FJ238471	26
mutS	2	AJ413210	23
mutS	3	AJ413208	23
mutS	4	AJ413209	23
mutS	5	JX876630	This study
recP	1	EU283342	26
recP	2	FJ238477	26
recP	3	FJ238476	26
recP	4	JX876631	This study
recP	5	JX876632	This study
recP	6	JX876633	This study
recP	7	JX876634	This study
xpt	1	FJ238483	26
xpt	2	FJ238486	26
xpt	3	FJ238482	26

TABLE S5.1. (*cont*)

Gene	Allele	GenBank acession number	Reference
xpt	4	JX876635	This study
xpt	5	JX876636	This study
xpt	6	JX876637	This study
xpt	7	JX876638	This study
xpt	8	JX876639	This study
xpt	9	JX876640	This study
yqiZ	1	FJ263616	26
yqiZ	2	FJ268647	26
yqiZ	3	FJ268654	26
yqiZ	4	FJ263615	26
yqiZ	5	JX876641	This study
emm	*stG1451*	KC291148	This study

TABLE S5.2. MLST data for 90 *S. canis* animal and human isolates from different geographic locations

Strain	ST	*gki*	*gtr*	*murI*	*mutS*	*recP*	*xpt*	*yqiZ*	Origin	Geographic Location
FMV1315.02	1	1	1	1	1	1	6	1	Dog	Portugal
FMV1425.02	1	1	1	1	1	1	6	1	Dog	Portugal
FMV2238.02	1	1	1	1	1	1	6	1	Dog	Portugal
FMV2322.02	1	1	1	1	1	1	6	1	Dog	Portugal
FMV334.03	1	1	1	1	1	1	6	1	Dog	Portugal
FMV3639.06	1	1	1	1	1	1	6	1	Dog	Portugal
FMV5002.06	1	1	1	1	1	1	6	1	Dog	Portugal
FMVES5.02	1	1	1	1	1	1	6	1	Dog	Portugal
FUB9212	1	1	1	1	1	1	6	1	Dog	Germany
FUB9324	1	1	1	1	1	1	6	1	Dog	Germany
FUB9383	1	1	1	1	1	1	6	1	Dog	Germany
FMV5857.06	2	2	2	2	2	2	1	2	Dog	Portugal
FUB23896	2	2	2	2	2	2	1	2	Dog	Germany
FUB24014	2	2	2	2	2	2	1	2	Dog	Germany
FMV565.02	3	3	3	3	3	1	2	3	Dog	Portugal
FUB24004	3	3	3	3	3	1	2	3	Dog	Germany
FUB24007	3	3	3	3	3	1	2	3	Dog	Germany
FMV3662.06	5	1	2	4	3	1	8	1	Dog	Portugal
FMV6216.05	6	2	2	2	2	2	4	2	Dog	Portugal
FUB23887	7	2	2	2	2	4	1	2	Dog	Germany
FUB6286	8	2	2	4	4	2	3	2	Dog	Germany
DSM20715^{T}	9	3	5	3	3	1	2	3	Cow	Reference Strain
FMV1434.06	9	3	5	3	3	1	2	3	Cat	Portugal
FMV2274.05	9	3	5	3	3	1	2	3	Dog	Portugal
FMV261.04	9	3	5	3	3	1	2	3	Cat	Portugal
FMV2797.03	9	3	5	3	3	1	2	3	Dog	Portugal
FMV2861.06	9	3	5	3	3	1	2	3	Dog	Portugal
FMV370.06	9	3	5	3	3	1	2	3	Dog	Portugal
FMV382.04	9	3	5	3	3	1	2	3	Dog	Portugal
FMV4037.03	9	3	5	3	3	1	2	3	Dog	Portugal
FMV5369.04	9	3	5	3	3	1	2	3	Dog	Portugal
FMV7307.07	9	3	5	3	3	1	2	3	Dog	Portugal
FUB11660	9	3	5	3	3	1	2	3	Cat	Germany
FUB11826	9	3	5	3	3	1	2	3	Cat	Germany
FUB12082	9	3	5	3	3	1	2	3	Cat	Germany
FUB12111	9	3	5	3	3	1	2	3	Cat	Germany
FUB23835	9	3	5	3	3	1	2	3	Dog	Germany
FUB24001	9	3	5	3	3	1	2	3	Dog	Germany
FUB5449	9	3	5	3	3	1	2	3	Dog	Germany
SH2062	9	3	5	3	3	1	2	3	Human	Portugal
SH2929	9	3	5	3	3	1	2	3	Human	Portugal
SH5085	9	3	5	3	3	1	2	3	Human	Portugal
SH5107	9	3	5	3	3	1	2	3	Human	Portugal
SH5276	9	3	5	3	3	1	2	3	Human	Portugal
FUB23886	10	3	5	3	3	1	9	3	Horse	Germany
FUB6202	11	4	2	4	4	2	3	2	Dog	Germany
FUB6204	11	4	2	4	4	2	3	2	Dog	Germany

TABLE S5.2. (*cont*)

Strain	ST	*gki*	*gtr*	*murI*	*mutS*	*recP*	*xpt*	*yqiZ*	Origin	Geographic Location
FMV5687.03	12	4	2	4	4	4	3	4	Dog	Portugal
SH5565	12	4	2	4	4	4	3	4	Human	Portugal
FMV1129.05	13	4	2	4	4	5	3	4	Dog	Portugal
FMV1565.02	13	4	2	4	4	5	3	4	Dog	Portugal
FMV183.06	13	4	2	4	4	5	3	4	Dog	Portugal
FMV1940.04	13	4	2	4	4	5	3	4	Dog	Portugal
FMV2014.02	13	4	2	4	4	5	3	4	Dog	Portugal
FMV2686.01	13	4	2	4	4	5	3	4	Dog	Portugal
FMV2863.03	13	4	2	4	4	5	3	4	Dog	Portugal
FMV2880.03	13	4	2	4	4	5	3	4	Dog	Portugal
FMV551.02	13	4	2	4	4	5	3	4	Dog	Portugal
FMV5829.03	13	4	2	4	4	5	3	4	Dog	Portugal
FUB6343	13	4	2	4	4	5	3	4	Dog	Germany
FUB9195	13	4	2	4	4	5	3	4	Dog	Germany
UNICAM15	13	4	2	4	4	5	3	4	Dog	Italy
FMV3142.03	14	4	2	4	4	6	3	4	Dog	Portugal
FMV3257.04	14	4	2	4	4	6	3	4	Dog	Portugal
FMV3818.05	14	4	2	4	4	6	3	4	Dog	Portugal
SH3379	14	4	2	4	4	6	3	4	Human	Portugal
FUB23675	15	4	4	5	4	4	7	1	Dog	Germany
FMV2460.07	16	4	5	3	4	1	2	3	Dog	Portugal
FMV596.02	16	4	5	3	4	1	2	3	Dog	Portugal
FUB11801	16	4	5	3	4	1	2	3	Cat	Germany
FUB23673	17	4	6	4	4	5	3	4	Dog	Germany
FUB23883	17	4	6	4	4	5	3	4	Dog	Germany
FMV1757.03	18	5	2	2	2	2	1	2	Dog	Portugal
FMV55.06	18	5	2	2	2	2	1	2	Dog	Portugal
FMV74.05	19	6	2	4	3	4	5	1	Dog	Portugal
FUB23971	19	6	2	4	3	4	5	1	Dog	Germany
FMV1451.06	20	7	4	5	3	1	7	3	Dog	Portugal
FMV4399.06	21	7	4	5	3	4	7	1	Dog	Portugal
FMV4489.05	21	7	4	5	3	4	7	1	Dog	Portugal
FMV6425.07	21	7	4	5	3	4	7	1	Dog	Portugal
FUB23999	21	7	4	5	3	4	7	1	Dog	Germany
FMV1916.05	22	8	2	3	3	1	2	4	Dog	Portugal
FMV3689.07	23	8	5	3	3	1	2	3	Dog	Portugal
FMV3835.03	23	8	5	3	3	1	2	3	Dog	Portugal
FUB23676	24	1	1	1	1	4	6	1	Dog	Germany
FUB12146	25	9	7	3	5	7	7	5	Cat	Germany
Sc2	2	2	2	2	2	2	1	2	Human	USA
Sc3	3	3	3	3	3	1	2	3	Human	USA
Sc4	4	4	2	4	4	3	3	4	Human	USA
FSL Z3-227	9	3	5	3	3	1	2	3	Cow	USA

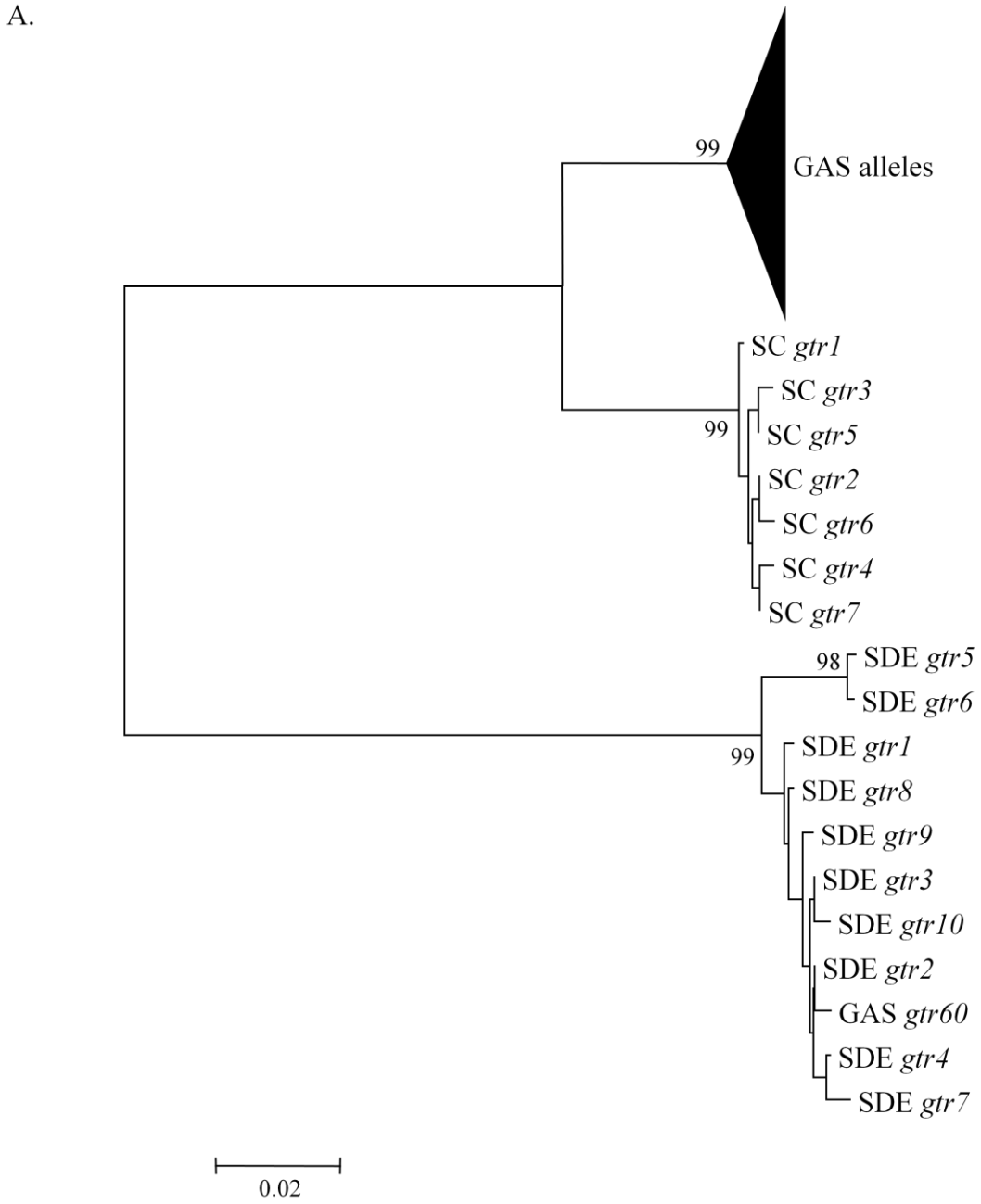

FIGURE S5.1. Neighbor-joining unrooted trees of *gtr* (A), *murI* (B), *mutS* (C) and *xpt* (D) alleles of *S. canis*, SDE and GAS

The trees were built with the methods indicated in materials and methods and in the legend of Fig. 1. *gtr* alleles are indicated as SC *gtr1* to SC *gtr7*, *murI* alleles as SC *murI1* to SC *murI5*, *mutS* alleles as SC *mutS1* to SC *mutS5* and *xpt* alleles as SC *xpt1* to SC *xpt9*. The sequences from the two comparator species were obtained from the respective databases (http://www.mlst.net/).

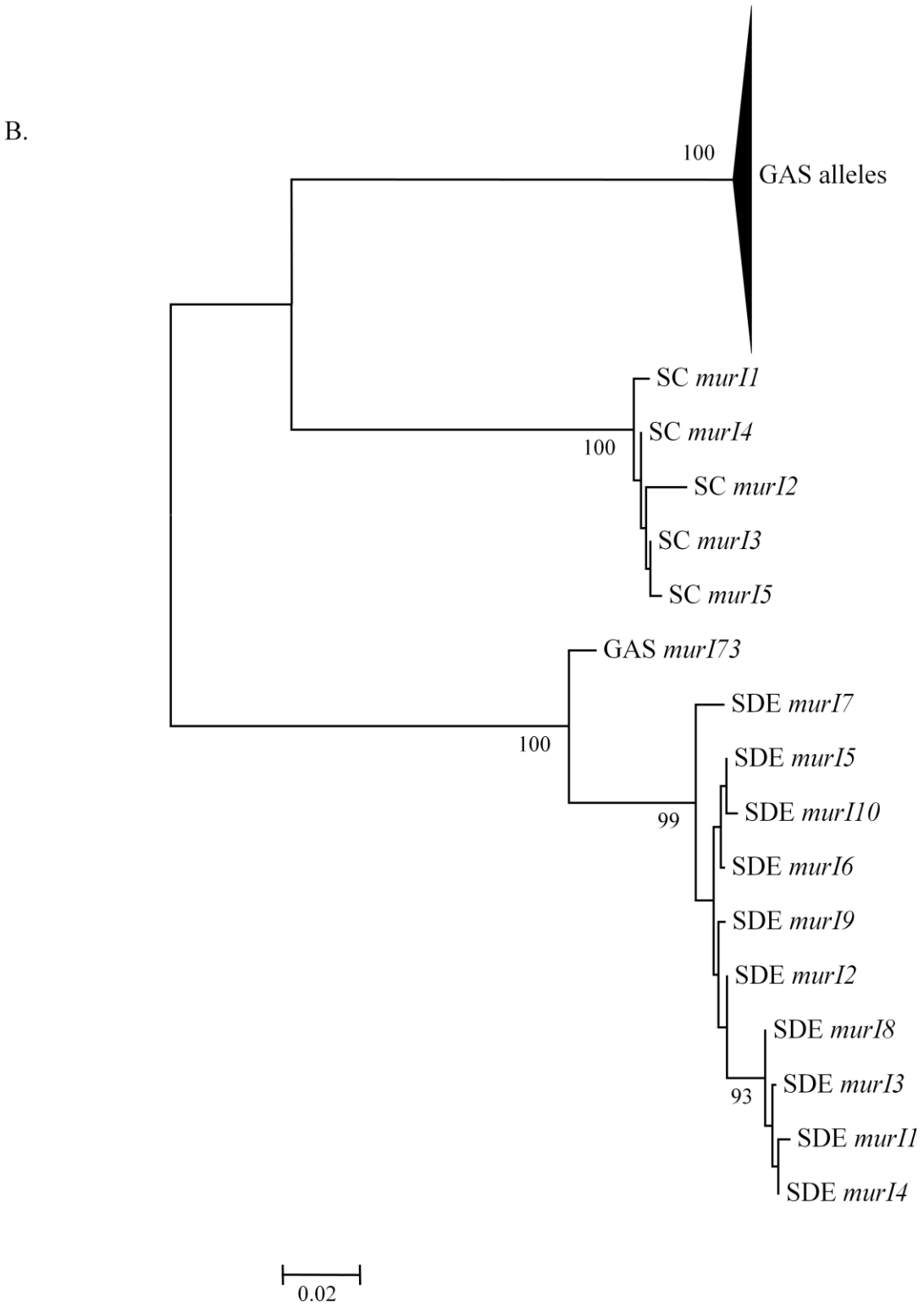

FIGURE S5.1. (*cont*)

C.

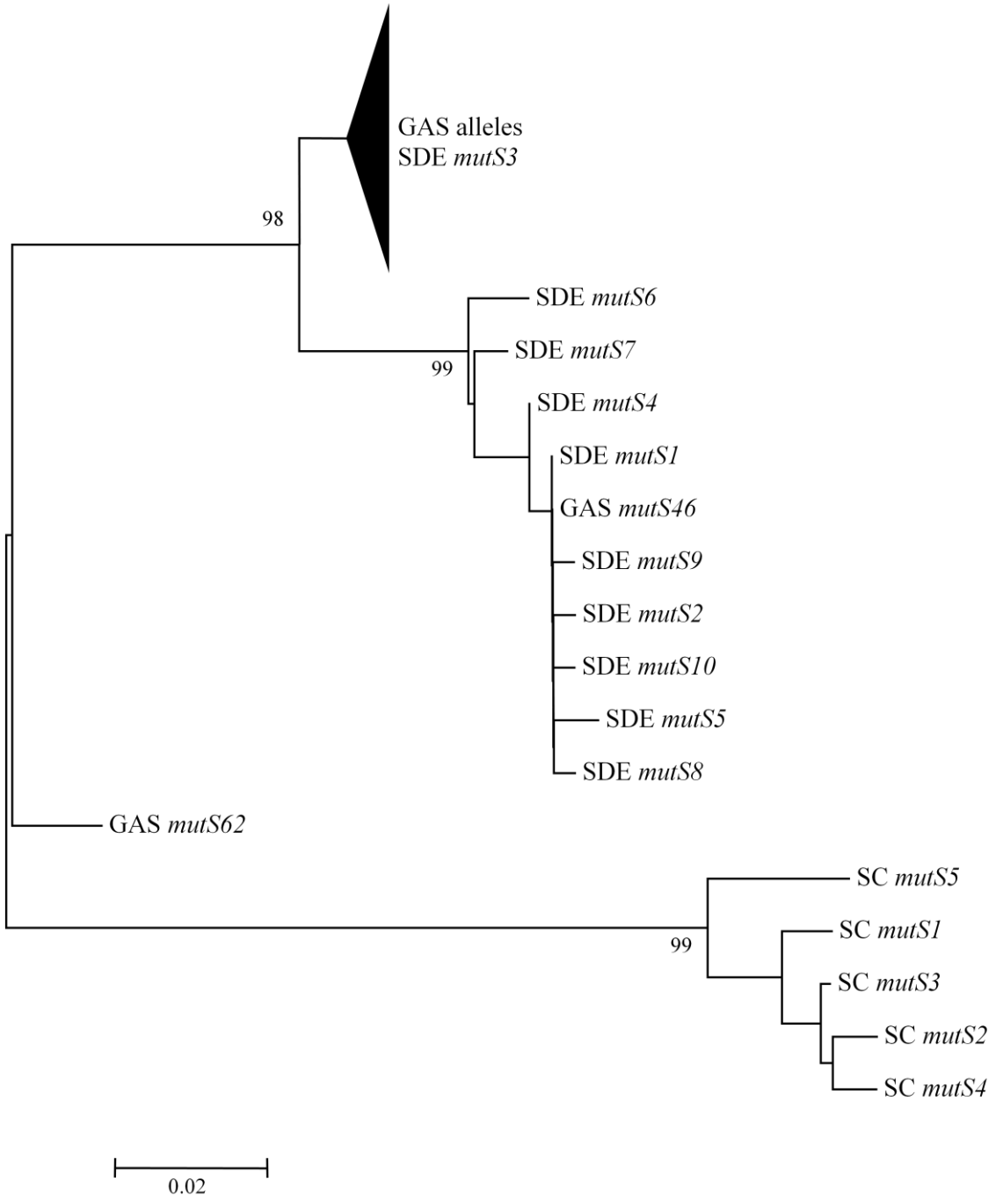

FIGURE S5.1. (*cont*)

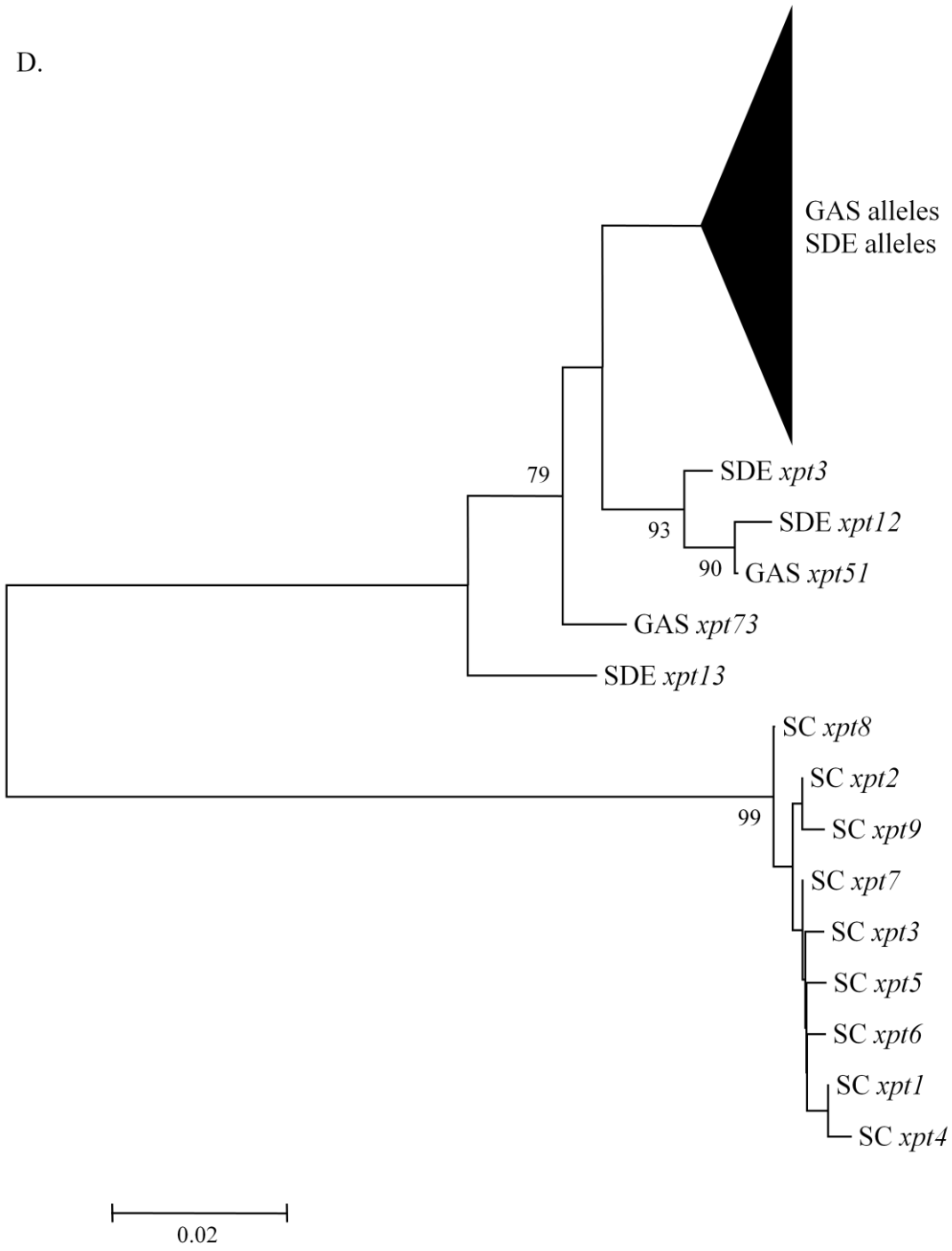

FIGURE S5.1. (*cont*)

CHAPTER 6

General Discussion

6. GENERAL DISCUSSION

The last decades witnessed an increased recognition of GCGS as human pathogens. This increment is particularly significant for SDE, given that a growing number of reports describing infections by this species are being published, many from cases of severe infection (1–3). Moreover, in some instances, SDE presents a disease burden which is close or even exceeds that attributable to GAS and GBS (4). Whether this is a true increase in the incidence of SDE infections or merely reflects a better awareness of its pathogenic role in the human host is a question not entirely solved, mainly because of the taxonomic issues that hampered many of the published GCGS literature and limit the assessment of the individual species contribution to the overall numbers reported. Nevertheless, a number of studies from distinct locations in the world showed increased incidence of GCGS infections over time (4–6), favoring the first scenario. Currently, SDE is no longer regarded simply as a member of the human microbiota showing low virulence to humans but as an emerging streptococcal pathogen (1, 7, 8).

A central question remaining to be answered is whether the increased detection of SDE has been driven by the emergence of more virulent strains of this species. Many authors speculated on the molecular events that could increase SDE virulence and searched for the genetic determinants that could mediate it. A common path was to look for genes known to encode proteins important for the pathogenesis of the closely related GAS (which is much better characterized), resulting in the description of a number of shared *loci* between the two species. In many instances such findings were interpreted as likely indication of the acquisition of genetic material through horizontal gene transfer events. The pathogenesis of SDE diseases is likely dependent on the action of many virulence factors present in the bacteria but host factors may also play a key role given the noted association of SDE disease with elderly patients and with pre-existing co-morbidities. However, in addition to studying the distribution of virulence genes throughout the SDE population, typing methods that allow an improved delineation of bacterial clones for this species are crucial in order to allow the identification of potentially more virulent strains that could have arisen during evolution.

Prior to the studies described in this thesis there was no information available on the epidemiology of GCGS infections in Portugal and molecular typing data from other geographic locations was also limited. This lack of both background information and

typing methods allowing the identification of GCGS clonal lineages prompted the development of the MLST schemes for SDE and *S. canis* described in this thesis. The data generated brought new insights into the population structure of these two species and also into the evolutionary dynamics which may contribute to part of their apparent increased success as human pathogens. The characterization undertaken also allowed some aspects related to virulence and antimicrobial resistance to be explored.

6.1. Issues relating to the identification of GCGS species

The importance of conducting species level identification of GCGS is today increasingly recognized and seen as critical to fully evaluate the prevalence, epidemiology and the pathogenic characteristics of each of the bacterial species included in the group. The identification procedures described in the current work showed that the identification of either SDE (chapter 2) or *S. canis* (chapter 5) do not pose critical difficulties. Both phenotypic methods, such as API 20 Strep used in this work or other biochemical tests, and genotypic methods (9, 10) allow the distinction between the two. The application of MLST, which generates sequence information for multiple *loci*, also allows unambiguous distinction of these species, since the vast majority of the alleles are species specific and the issues related to single *locus* based identification, such as low discriminatory separation or the sharing of alleles by distinct species, are overcome.

The current work showed the dominant role of SDE among GCGS species causing human infections in Portugal, as seen in many other geographical regions (2, 5). SDE was the only species isolated in the first years of the study and only when the collection was expanded by enrolling a growing number of laboratories were *S. canis* isolates recovered. Thus, it is conceivable that in the beginning of the study the sample size was not large enough to allow detection of *S. canis*, given the low prevalence of this species compared to SDE. Studies conducted elsewhere involving a low number of isolates may have also a limited ability to detect the presence of *S. canis* or other GCGS species which behave as infrequent human pathogens.

The characterization of SDE isolates conducted in chapters 2 and 4 showed that isolates of this species bearing either the Lancefield group C or group G antigens were not clearly separated from each other. Not only did these serological distinct isolates have common 16S rRNA alleles and biochemical profiles, but they were also not

separated by PFGE and shared some *emm* types, a finding reported elsewhere (11). Such observations were in agreement with previous studies that supported their inclusion under a common epithet (9, 12). The application of MLST described in chapter 3 further shows that Lancefield group carbohydrates are associated with distinct SDE lineages, given that a good correlation was found between both MLST STs and CCs with Lancefield classification, although no long-term genetic isolation between them could be found. Thus, the data generated in this thesis highlight the absence of genetic evidence for the separation of SDE commonly found in human infections bearing these groups and reinforces the importance of studying strains bearing both groups together to fully evaluate the significance of SDE as an agent of human infection. Studies that characterize exclusively Lancefield group C or G strains, a concept still in use in the study of GCGS infections, may suffer from a double drawback: not only some GCGS species may be missed, but also, in the particular case of SDE, they have the bias of selecting distinct clones, resulting in an inadequate picture of the SDE population structure.

6.2. The *emm* gene: epidemiologic and virulence considerations

The *emm* typing technique has gained acceptance in SDE epidemiology during the last two decades and has now become the most widely used typing technique for this species. Previous studies have described the *emm* type distribution among isolates recovered mostly from invasive infections (13), but the work described in Chapter 2 was the first trying to correlate *emm* types with virulence in SDE, *i.e.*, to look for *emm* type differences in SDE isolated from invasive and non-invasive infections. This approach allowed the identification of two *emm* types, *stG2078* and *stG10*, both statistically overrepresented among invasive isolates in the sampled population, suggesting an association between *emm* type and invasive disease potential. The complementary PFGE analysis showed the polyclonal origin of *stG2078* isolates (further suggesting the importance of this particular *emm* type or of a closely linked determinant for virulence), while the nonrandom clustering of *stG10* isolates in a single clone suggested the contribution of a common genetic background.

emm types *stG2078* and *stG10* are well represented among studies characterizing SDE invasive isolates published in the last years (Fig. 6.1), although this observation cannot be generalized to all geographic locations (1, 14). Interestingly, both were found

among a group of uncommon *emm* types associated with more severe disease and increased mortality rates in patients with SDE bacteremia in Finland (1). A study from Japan also characterizing SDE isolates from both invasive and non-invasive infections found *stG485*, *stG6792* and *stG2078* significantly associated with invasive disease (15), supporting our own findings on the increased virulence of strains presenting this last *emm* type. Nonrandom distributions of *emm* types have occasionally been noted in other studies, as was the case of *stC36* and *stC839* detected only in throat isolates and s*tG480* and *stG6* only in blood isolates in a Danish hospital (16). All these observations support a different virulence of SDE strains of distinct *emm* types, although it is not clear whether such differences are driven by the direct action of the product of the *emm* gene, a closely linked gene or other factors in the genome of a genetic lineage exhibiting a given *emm* type.

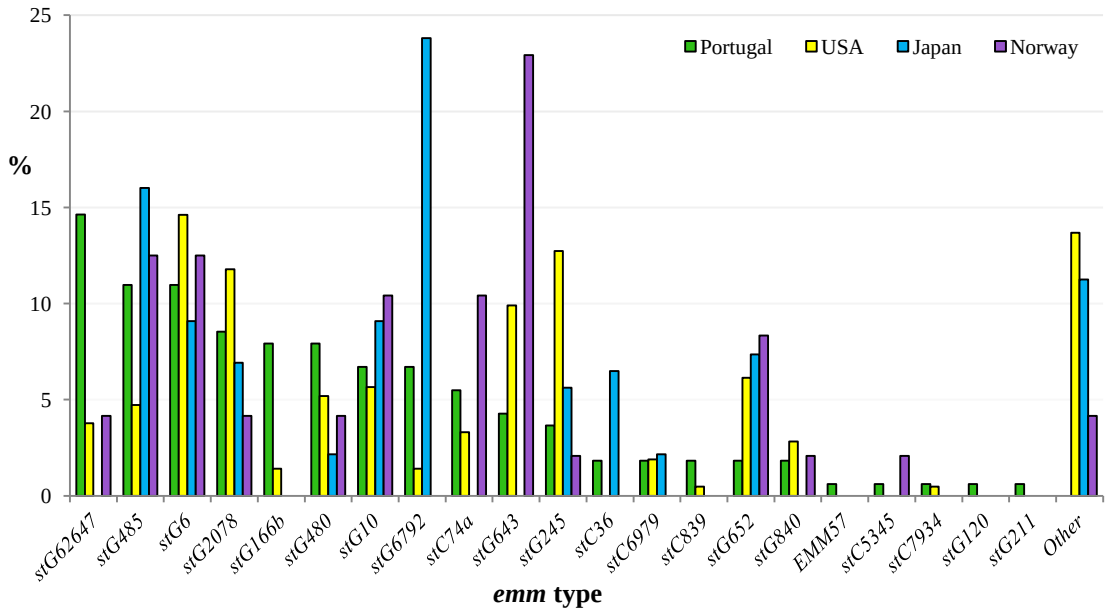

FIGURE 6.1. *emm* types distribution among SDE invasive isolates in selected studies from distinct geographic locations

emm types are given by decreasing order of frequency as they were found in 164 isolates recovered in Portugal from 1998 to 2010 [work not published in scientific paper (17)]. Isolates used for comparison include 212 isolates recovered in USA from 2002 to 2004 (2), 231 isolates recovered in Japan from 2006 to 2007 (4) and 48 isolates recovered in Norway from 2006 to 2009 (18). The "other" *emm* types column includes more than 15 different *emm* types, none including more than 10 isolates.

The several studies on SDE *emm* type distribution that are now available from distinct regions in the world show that the prevalence of individual *emm* types varies

geographically. Taking as example the studies referenced in Fig. 6.1, it can be seen that although a common group of *emm* types account for the majority of SDE infections independently of the geographic region considered, the individual *emm* type most prevalent at each geographic location changes. In some instances, *emm* types highly prevalent in a given geographic location are absent from another, as is the case of *stG62647* which was the most frequent in Portugal (when all isolates recovered from 1998 to 2010 are considered) (17) but not found in Japan (4) and with low prevalence in other locations. On the other hand, a number of studies have confirmed the importance of SDE isolates bearing *emm* type *stG6792* in Japan. In this country, *stG6792* isolates were shown to have a nationwide distribution (4), were associated with invasive disease (15) and poor outcome (4), and often identified among SDE isolates causing STSS (19). This *emm* type has much less expression among invasive infections in Portugal and was exclusively found among non-invasive isolates prior to 2006 (chapter 2).

There is still no study covering a sufficiently wide time range to provide evidence of *emm* type temporal variation. However, it has been observed that the *emm* types associated with SDE isolates causing STSS and necrotizing fasciitis in Japan have changed during time (3). Both geographic and temporal variations make it difficult to evaluate the association of individual *emm* types to virulence. It also implies that *emm* type differences observed on selected geographic locations and over restricted time periods should be interpreted with caution and that local dynamics should be taken into consideration.

A few aspects should be considered which point towards the importance of the genetic background of SDE isolates in virulence rather than the *emm* gene itself: (i) the *emm* type variability observed among invasive SDE isolates (2, 4, 13), an indication that multiple variants of this protein can be responsible for invasive infections; (ii) *emm* types commonly found among invasive isolates, even those for which occasional statistic association with invasive disease has been obtained, are not uncommon among non-invasive isolates, as is the case of the *emm* types referred above; (iii) the geographic differences found in the prevalence of different *emm* types among invasive and non-invasive isolates which may imply the presence of distinct clonal lineages associated with the same *emm* type. The application of complementary typing methods should help to clarify these issues.

Expanding the analysis beyond the *emm locus* should prove useful in defining which other factors contribute to the increased virulence of particular strains. Characterizing the region surrounding the *emm* gene in SDE might clarify if there are closely located gene(s) that could potentially influence virulence. As the *emm* gene is located in a distinct genomic environment in SDE and GAS (20), the importance of this region to virulence in SDE cannot be inferred from the previous observations in GAS. Not only is evidence lacking that the *emm* gene is regulated in a similar manner in both species, but a recent study showed that the two types of multigene regulator like regions found in SDE are correlated with *emm* types but not with the site of infection, whereas in GAS the presence of *emm*-like genes in this region was associated with tissue tropism and disease presentation (14).

The contribution of other factors present across SDE genomes should also be clarified. In some cases, there is the need to elucidate how a given protein functions in SDE, since the knowledge on the mode of action for many of the presumed virulence factors comes from studies on GAS and the SDE counterpart may present unique features. On the other hand, studies exploring a more global view of the bacterial gene content should elucidate how different *loci* contribute to the overall virulence of this species and how they are controlled and interact with each other. The distribution of such factors in SDE population should be determined, as their presence/absence may explain the differences observed between strains. For example, a recent study detected a streptococcal inhibitor of complement-mediated cell lysis (SIC)-like gene in *stG2078* isolates (21), which has restricted distribution in both SDE and GAS.

6.3. Improving GCGS typing by applying MLST

Although *emm* typing is the only technique employed to type GCGS isolates in many studies, it is now becoming apparent that it is insufficient to ascertain the genetic relatedness between these strains. As shown by the simultaneous use of distinct typing techniques, not only does *emm* typing produces stratifications which are discordant with those produced by either PFGE or MLST for SDE, it is also unsuitable for this purpose in other GCGS species, as is the case of *S. canis*, since only a few strains will amplify a PCR product by the technique usually employed. Studies that complement the data obtained by *emm* typing are still scarce, and further work that does not rely solely on this technique should prove useful in defining the population dynamics of GCGS

species, its temporal evolution, and in providing a framework upon which the question of whether subpopulations of SDE strains have enhanced virulence can be addressed.

Given the established usefulness of the MLST technique for molecular typing of many bacterial species, new MLST schemes for SDE and *S. canis* were developed in collaboration with other groups. The work reported in this thesis occurred simultaneously and was preceded by the study published by Ahmad and colleagues in 2009 (11), which characterized isolates from both species recovered from invasive infections in the United States. In the case of SDE, we developed an independent MLST scheme with alternative primer pairs but targeting the same regions of the genome. We provided a large new dataset covering isolates from distinct geographic locations recovered from invasive and non-invasive infections, which improved our understanding on SDE population structure and highlighted the broad geographic distribution of STs. For *S. canis*, the proposed modifications resulted in a fully applicable MLST scheme allowing the characterization of a large number of *S. canis* isolates from human and animal sources. This work also aimed at establishing a publicly accessible MLST database for each of these two species (http://sdse.mlst.net/ for SDE and http://pubmlst.org/scanis/ for *S. canis*), with the goal of improving the utility of MLST by allowing the information obtained by distinct laboratories to be easily shared.

A high level of genotypic diversity was found among SDE isolates analyzed from Portugal, supported by the variety of *emm* types, PFGE clusters and MLST defined STs that were found. Similarly, the analysis of *S. canis* isolates also pointed to an heterogeneous population in which, despite their low number, human isolates had STs which were representative of the several lineages found in animal infections. The polyclonal nature of both SDE and *S. canis* shows that the ability to cause disease in humans is not a particular characteristic of a few GCGS genetic lineages. While this is not surprising for SDE, since it is well known to be colonizing the human host, the diversity of *S. canis* isolates causing human infection suggest that factors distributed in the genome of numerous genetic lineages of this species allow them to infect humans once there is an opportunity of transmission. Whether those factors also allow *S. canis* to persist or colonize the human host remains to be demonstrated, but these observations suggest a scenario in which these GCGS species have vertically inherited characteristics that allow them to infect the human host. As shown by the recently published genomes

of both SDE and *S. canis* strains (22, 23), multiple factors are present in both these species that are shared with the human pathogen GAS, which may contribute to their virulence in the human host.

The work conducted in this thesis also provided evidence for the clonal expansion of some specific genetic lineages. In the case of SDE, this was supported by the increased prevalence of some *emm* types and PFGE clones when these two typing techniques were employed (e.g. the *stG10* invasive isolates in chapter 2 and the *stG166b* levofloxacin resistant isolates in chapter 4). The MLST analysis of isolates recovered in distinct geographic locations was instrumental in showing the predominance of a few SDE clones among the global SDE population (Fig 3.1, chapter 3) and a similar situation could be observed with the predominance of ST9 among *S. canis* isolates from different countries (chapter 5). Despite the clonal diversity observed at regional level, the demonstration that a small number of STs are highly prevalent and have a wide geographic distribution shows the presence of successful genetic lineages which may differ from the remaining population in terms of increased fitness, virulence properties or transmission capacity. Further characterization should unveil whether this is the case and if these properties are linked.

The SDE MLST analysis performed in chapter 3 included isolates derived from large independent collections, with the aim of compiling a genetically diverse dataset to allow picturing the global population structure of this bacterium. The sampling strategy employed is limited in the sense that isolates were chosen based on prior knowledge of their *emm* type and geographic site of isolation. Thus, it does not necessarily reflect clonal abundance nor the clonal structure of the population at each of the geographic locations included in the study. MLST application to SDE populations isolated from specific geographic locations should clarify the most important genetic lineages circulating in that area and allow the regional population structure of SDE to be better clarified.

The application of MLST to SDE population isolated in Portugal is still partial, as only 36 isolates were included in the analysis conducted in chapter 3. Since this work, further characterization on SDE invasive isolates recovered in Portugal was conducted (17) (*emm* type distribution is shown in Fig. 6.1) and a subsequent study typed a collection of SDE isolates mainly recovered from the pharynx of Indian children (24), reporting for the first time the MLST characterization of asymptomatic carriage isolates.

By including this data, Fig. 6.2 shows all the isolates characterized to date by this MLST scheme and intends to illustrate what MLST has brought to the current knowledge on the population structure of SDE. The presence of dominating SDE genetic lineages with a worldwide distribution as described in chapter 3, can be visualized in the CCs represented in Fig. 6.2, most of them including isolates recovered from invasive and non-invasive infections and colonization, recovered in distinct locations. The same holds true for the putative founders of these CCs. However, this set of SDE isolates, enriched in invasive isolates from Portugal and non-invasive/colonizing isolates from India, also shows the presence of STs or CCs which are exclusively found in these regions, likely representing lineages that are particularly prevalent in these geographic locations. The importance of local dynamics is also exemplified by the findings on levofloxacin resistance reported in this thesis. High rates of resistance to levofloxacin are still observable in Portugal at present (unpublished data), whereas resistance to this agent has not been found in studies conducted worldwide (25), indicating that local factors can be contributing to the spread of particular resistance determinants.

The study on the MLST characterization of other important GCGS species in human infections, *S. canis*, directly compared isolates from human and animal infections at the population level. Not only was the ability of several *S. canis* lineages to infect humans established but it was also shown that all *S. canis* human isolates characterized to date by MLST belong to the most prevailing lineages found in house pets. It is conceivable that being more frequent in the animal population, these lineages may have more chance to pass to humans. A recent study by Richards *et al* (23) applied a MLST scheme using a distinct set of housekeeping genes to type bovine and house pet isolates recovered in United States and Europe. The genome sequenced *S. canis* strain was found to have the predominant ST in that study, which occurred in different hosts and geographic locations, an observation identical to the one reported in this thesis. Although not directly comparable, this is a strong indication that the results of both studies are identical, and confirm that the same *S. canis* lineages are found in bovine, house pet, and human hosts.

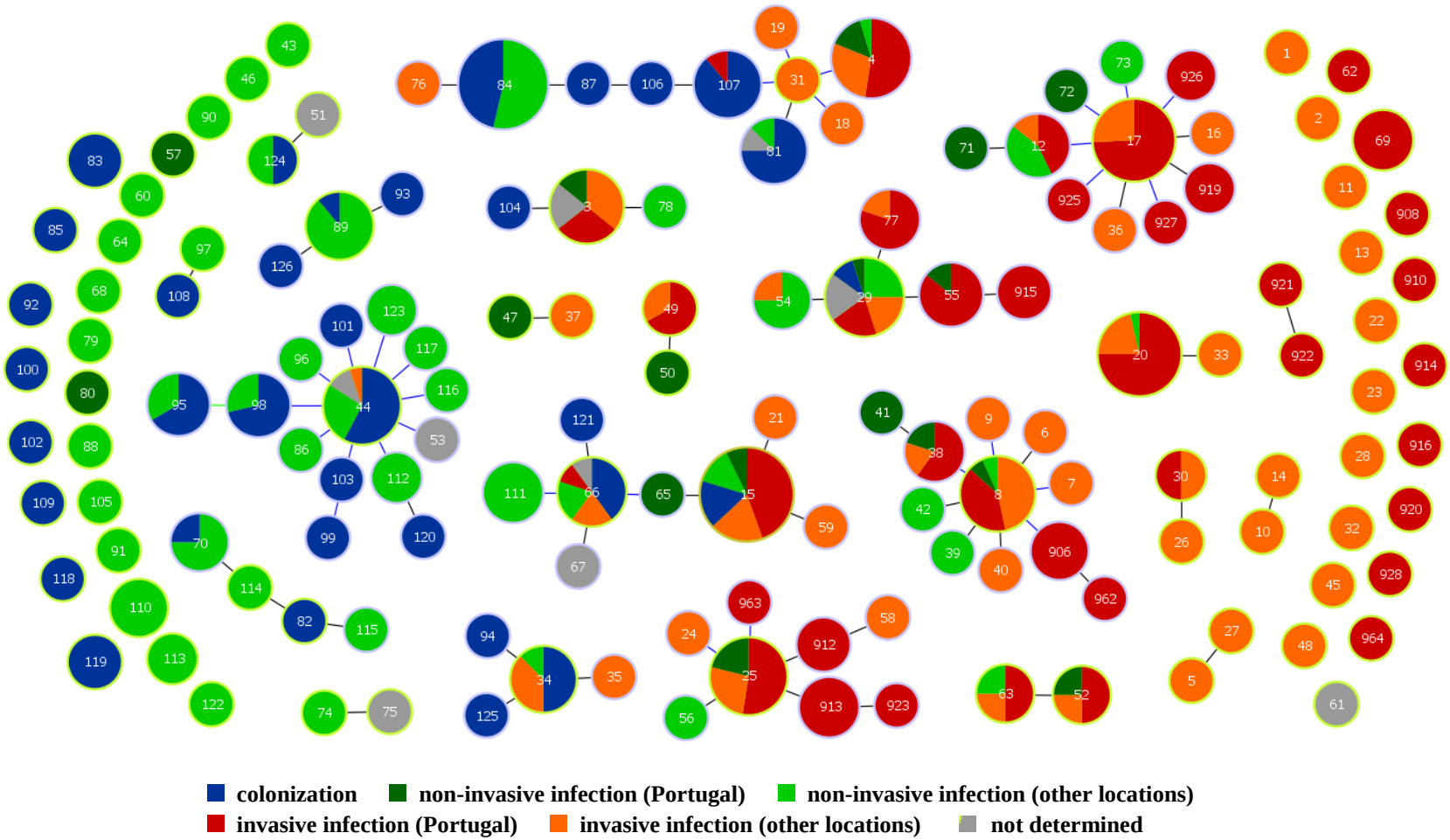

FIGURE 6.2. goeBURST diagram of relationships between 513 SDE isolates

The size of each circle is proportional to the number of isolates with that particular ST on a logarithmic scale. STs assigned to the same CC at the SLV level are linked by straight lines. Putative CC founders are identified by a light green line.MLST data used to construct the diagram include the reported in chapter 3, complemented with all invasive isolates recovered in Portugal until 2010 [work not published in scientific paper (17)] and from an Indian study including non-invasive and carrier isolates (24).

6.4. The relevance of horizontal gene transfer events in GCGS evolution

A noteworthy aspect of the work presented in this thesis was the evidence obtained for the occurrence of horizontal gene transfer events involving SDE and, to a lesser extent, *S. canis*. Several examples suggestive of recombination were found in genes encoding essential functions in the bacterial cell, including the DNA topoisomerase IV *parC* gene and the housekeeping genes included in the MLST schemes. Although this was not the primary goal and the approach used targeted the sequence of DNA regions accounting for an extremely small portion of the bacterial genome, the detection of such events indicates these must be a frequent phenomenon with a potentially crucial role in GCGS biology and evolution.

The importance of these events for GCGS has deserved special attention because of the possible acquisition of virulence genes from more pathogenic species, resulting in either an enhancement of the invasive capacity of GCGS species towards humans, such as SDE, or enabling the occupation of novel niches by zoonotic pathogens. This idea is implicit for many of the GAS virulence genes which have been detected in SDE, and is supported in some cases by their observed association with mobile genetic elements. Most lines of evidence point towards the occurrence of genetic changes between SDE and GAS. Examples involving the two species are countless and include genes encoding the M protein (26), a fibronectin-binding protein (27), or the streptokinase gene (28). The MLST data presented in this thesis showing horizontal gene transfer involving the housekeeping alleles of these two species is in accordance with contemporary or subsequent studies employing this technique (11, 24). Moreover, the recently available full genome sequence data for SDE showed abundant horizontal gene transfer between the core genomes of the two species with more evidence for gene flow in the GAS to SDE direction (29).

The relevance of such events is also exemplified in the case of the *emm* gene, which assumes importance by its potential direct impact in the virulence. The transfer of *emm* genes has been assumed for years (26) and *emm* types initially found in GAS, such as *emm57* and *emm12*, have long been detected in SDE (30). SDE alleles were also occasionally found in GAS, as was the case of *stG1750* detected in invasive GAS isolates from Portugal (31) and *stCK401* in GAS from Australia (32). However, such examples seem to be exceptions, as alleles found in each species are almost exclusive as is shown by the *emm* type analysis described in this thesis and supported by the vast

majority of studies performing *emm* typing in SDE (33, 34). The existence of intraspecific transfers in SDE is supported by the MLST analysis performed here and by others (11), which found the same *emm* type in distantly related STs and also distinct *emm* types in a single ST. One of the likely explanations for this lack of congruence between *emm* type and SDE clonal lineages, also observed in PFGE (chapter 2), is the existence of recombinational replacements involving the *emm* gene. Evidence for the occurrence of such events was produced by a Norwegian study (34) which demonstrated some SDE isolates possessed *emm* genes composed of segments derived from distinct SDE associated *emm* types. New genomic information which identifies the *emm* gene as part of an ancient pathogenicity island shared with GAS but disaggregated in SDE (35, 36) reinforces the role that horizontal gene transfer and recombination may have had in the evolution of this major virulence factor.

Despite the significance of the exchanges between SDE and GAS, the horizontal genetic transfer events occurring including GCGS have much of partners. SDE gene transfer with GBS has been documented (37–39) and includes examples such as the C5a peptidase or the surface protein Lmb, whose encoding genes are found in a putative composite transposon flanked by insertion sequences in both species (39). SDE transfers with other GCGS species have also been reported, as was the case of the detected divergent MLST alleles for *S. canis* which indicated a gene flow in the direction SDE to *S. canis*. Such events were recently detected in the published genome of *S. canis*, involving *S. agalactiae* and *S. dysgalactiae* subsp. *dysgalactiae* (23). GAS has also been implicated in the occurrence of horizontal gene transfer with other GCGS species. An example is the presence of similar exotoxin genes in GAS and GCGS species which do not infect humans such as *S. dysgalactiae* subsp. *dysgalactiae* (40) or *S. equi* subsp. *equi* (41), in spite of the fact that each of these species behave as host specific pathogens and are not recognized to share the same host. Thus, a growing body of evidence is being generated for the existence of a common gene pool between GCGS species and other beta-hemolytic species, in which in addition to the common evolutionary history of these bacteria, horizontal gene transfer events are likely to contribute significantly to the genetic diversification of individual strains.

The last years have also been fruitful in clarifying the possible mechanisms of horizontal gene transfer occurring in GCGS. Several mobile genetic elements are now known to be present in the genomes of these species which could mediate such events,

and include bacteriophages, plasmids or integrative conjugative elements (22, 23). It has long been known that some bacteriophages could propagate, lyse and lysogenize both GCGS and GAS (42) and to transduce streptomycin resistance markers from group A to group G streptococci (43). More recently, prophages significantly similar to GAS prophages were identified in SDE (22, 44), *S. equi* subsp. *equi* (45) and *S. dysgalactiae* subsp. *dysgalactiae* (36). A rising number of integrative conjugative elements has been described in SDE that were shown to carry genes presumably involved in virulence and resistance to metals (46) or to encode putative secreted extracellular proteins (38). Suzuki and colleagues (36), in their whole-genome comparison of *S. dysgalactiae* subspecies and GAS, identified a predominance of transposons and transposase genes which ranked them among the top of streptococcal genomes in terms of transposase abundance. A putative integrative plasmid was also recently identified in the *S. canis* genome that was shared with GBS (23). Distinct elements may mediate transfer between distinct species, and one study indicated that SDE bacteriophages might play a major role in transfers between SDE and GAS while SDE exchanges with GBS would be predominantly mediated by transposons (37). The chimeric nature evidenced by many of these elements (22, 44, 47–49) also highlights the role of recombination in their assembly and gene content.

7. CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The work reported in this thesis provides novel insights into the biology of GCGS in relation to the human host, with special focus on aspects related to population dynamics, virulence and antimicrobial resistance. The molecular epidemiology and evolutionary relationships among members of the two GCGS species studied, SDE and *S. canis*, were scrutinized at both regional and global levels. The clonal relationships among strains were elucidated and a polyclonal nature of SDE and *S. canis* populations revealed, supported by the identification of a great diversity of genetic lineages shown by different typing methods. The pathogenic potential of multiple members of these two species was substantiated by the observed ability of several strains to cause disease in the human host.

The use of the *emm* gene as an epidemiologic marker was shown to be complemented by applying other techniques, as *emm* typing produced stratifications that were not in agreement with either PFGE or MLST. The establishment of MLST schemes for SDE and *S. canis* provides new tools for future studies on their population dynamics. Relevant findings in our work also included the statistic association of certain *emm* types with invasive SDE isolates, suggesting a role for the M protein in SDE virulence. The data also produced evidence for recombinational replacements between SDE and GAS in several *loci* and, to a smaller extent, between SDE and *S. canis*, indicating that horizontal gene transfer events are important mechanisms driving genetic variability in GCGS populations, which may impact key bacterial functions such as virulence and antimicrobial resistance.

The novel data available, mainly from whole-genome sequencing projects, has confirmed that these bacteria have the gene repertoire consistent with their pathogenic role and potential to infect the human host. Thus, continuous surveillance that will contribute to elucidating their pathogenic potential and identifying the factors contributing to their virulence and dissemination, both at the strain and population levels, are needed.

In view of the work presented in this thesis and the current knowledge on GCGS pathogenesis in the human host, key issues remain to be elucidated. Thus, among the possible aspects which may be assessed in future work, the following can be enumerated:

- Monitor the GCGS species causing infections in humans in Portugal. The continuous characterization of these isolates will allow determining whether there is an increase in the incidence in human infection of any of the GCGS species and if the epidemiology reported in this thesis is maintained.

- Continue the molecular characterization of GCGS isolates, in order to obtain a better recognition of the most significant genetic lineages causing human infections. Detailed typing characterization covering a longer time span may provide additional information on epidemiologic temporal trends and help to understand the population dynamics of these pathogens.

- Determine the most suitable typing method(s) to be used for the characterization of GCGS species. The quantitative assessment of the congruence obtained by employing distinct typing methods to these bacteria has been limited. An approach centering on evaluating the discriminatory ability of each typing method, together with the analysis of the correlation of the different techniques, may help improving the delineation of GCGS clonal lineages.

- Expand the observation of an association between *emm* type and virulence in SDE by exploring features of either the *emm* gene or of the genomic environment surrounding this *locus* that may impact on the virulence of the bacteria. The study of isolates with *emm* types associated with invasive disease may be of particular interest.

    - Explore intrinsic features of the *emm* gene: the *emm* type is defined by the analysis of a small proportion of the *emm* gene. Differences in the biological properties of M proteins exhibited by SDE isolates may be encoded in regions located outside the region analyzed by the *emm* typing technique. Sequence analysis of the entire length of the *emm* gene in SDE will allow the comparison of the composition of the distinct proteins encoded and may reveal differences within the same *emm* type.

- Explore differences in the genomic environment surrounding the *emm locus*: the availability of SDE genomes allows the genomic context to be elucidated to a greater extent. Given that the origin of the *emm locus* in SDE may be related to an ancient pathogenicity island shared with GAS, specific studies can test the possibility of some *emm* regions of SDE having an organization similar to that of GAS. Whether specific *emm* genes may be transferred between SDE strains as part of one element of this kind may also be questioned.

- Characterize the *locus* of the *emm* types detected among the *S. canis* isolates analyzed in this thesis. Although in the *S. canis* published genome a few *emm*-like genes were present (23), none was identical to the *emm* types previously described. Thus, it would be of interest to define the exact *locus* from which *emm*-like genes are being amplified in *S. canis* isolates positive in the *emm* typing scheme, since these could represent *loci* that are more related to both SDE and GAS *loci.*

- Determine the distribution of genes encoding putative virulence factors which are differentially distributed throughout the GCGS populations and may account for differences in strain invasiveness.

    - Investigate the presence of genes encoding pilus-like structures in the FCT-like regions detected in SDE. Given the major contribution that pili are thought to have to the ability of bacteria colonizing and infecting their hosts, including the beta-hemolytic GAS and GBS, these factors are likely to have an important role in GCGS as well.

    - Search for the presence of superantigen genes in GCGS. Despite the lack of evidence for the expression of the *speG* gene in SDE, and the absence of other superantigen genes in most of the SDE isolates analyzed to date, distinct approaches may still be considered. For example: the design of new primers which take into account the allelic variation observed in some of these genes (including the allelic variants observed in other GCGS species); screening for the genes recently described in *S. equi* subsp. *zooepidemicus* (50); or extend the analysis to other GCGS species (where the products of these genes may have increased relevance). Studies that involve these

approaches complemented with expression analysis may generate important new information.

- Define the taxonomic status of SDE from animal origin, namely the beta-hemolytic Lancefield group C and L strains commonly isolated from house pets, cattle and other animals. As it is still a matter of debate the classification of these strains as subspecies *equisimilis* or *dysgalactiae* (16), work should be done to clarify their taxonomy and their relative clinical importance to the various hosts from which they have been isolated.

- Clarify the occurrence of SDE strains from animal origin in humans. The comparison of beta-hemolytic SDE from human and non-human sources should elucidate the origin of group L strains detected in both animals and humans. For example, application of the MLST technique to characterize animal isolates of SDE (as illustrated by the work on *S. canis* reported in chapter 5 of this thesis), may allow the identification of SDE strains isolated from human infection that are of zoonotic origin, determine how frequently they occur, and improve their detection.

- Conduct studies on carriage of GCGS. Most of the available studies were conducted prior to the current species delineation implying that speciation was not performed or is incomplete. New carriage studies may:
    - Offer new insights into the frequency of SDE in colonization.
    - Allow the comparison to isolates causing infection. Characterizing the strains that are found in the same time frame in colonization and infection in a given geographic region, may be particularly useful in further defining if there are clonal lineages more prone to cause disease.
    - Test for the occurrence of colonization by other GCGS species.
- Perform whole-genome analysis of SDE and *S. canis* isolates. Given the impact that this technology is having in elucidating the molecular mechanisms driving GCGS evolution and virulence, it would be of interest to extend the number of isolates analyzed. In particular, new data could be obtained by:
    - Analyzing SDE isolates of the *emm* types that were associated with invasiveness.

- Sequencing SDE isolates from animal origin, in order to determine how genetically and pathogenically distinct are these strains from human SDE. At the current time, SDE isolates of human origin have been compared to a *S. dysgalactiae* subsp. *dysgalactiae* isolate for which genome information is also available (36). To extend this analysis to SDE isolates from distinct hosts may help bring novel insights into the mechanisms associated with niche separation and those important for pathogenesis in the human host.
- Characterizing additional *S. canis* isolates. A single *S. canis* genome has been published (23), corresponding to an isolate from bovine mastitis. By sequencing other isolates, the virulence factors present in this species could be explored, and the inclusion of human isolates could further test whether the presence in the human host is associated with any specific traits in individual strains.

• Define the resistance mechanisms to different antimicrobial classes in SDE. Monitoring resistance to agents which are treatment options for the management of GCGS infections is critical as significant resistance rates have been documented among streptococci and common resistance gene pools are present. The follow up of the high resistance rate to fluoroquinolones detected should also be maintained, as the observations reported for SDE isolated from human infections in Portugal remain unique at the present time.

The work presented in this thesis provided novel insights into the population structure of both SDE and *S. canis*. Integrated typing information together with new genomic data from genome sequencing projects should prove instrumental in elucidating the pathogenic potential of these GCGS. The availability of new GCGS genomes will allow to move into the field of comparative genomics that may allow the identification of the core and accessory genomes, and to infer more accurately the impact of intra- and inter-species transfer of genetic material in GCGS biology and evolution.

8. REFERENCES

1. **Rantala S, Vahakuopus S, Vuopio-Varkila J, Vuento R, Syrjanen J**. 2010. *Streptococcus dysgalactiae* subsp. *equisimilis* bacteremia, Finland, 1995-2004. Emerging Infect. Dis. **16**:843–846.

2. **Broyles LN, Van Beneden C, Beall B, Facklam R, Shewmaker PL, Malpiedi P, Daily P, Reingold A, Farley MM**. 2009. Population-based study of invasive disease due to beta-hemolytic streptococci of groups other than A and B. Clin. Infect. Dis. **48**:706–712.

3. **Ikebe T, Oguro Y, Ogata K, Katsukawa C, Isobe J, Shima T, Suzuki R, Ohya H, Tominaga K, Okuno R, Uchitani Y, Tada Y, Okabe N, Watanabe H**. 2010. Surveillance of severe invasive group G streptococcal infections in Japan during 2002-2008. Jpn. J. Infect. Dis. **63**:372–375.

4. **Takahashi T, Sunaoshi K, Sunakawa K, Fujishima S, Watanabe H, Ubukata K**. 2009. Clinical aspects of invasive infections with *Streptococcus dysgalactiae* ssp. *equisimilis* in Japan: differences with respect to *Streptococcus pyogenes* and *Streptococcus agalactiae* infections. Clin. Microbiol. Infect. **16**:1097-1103.

5. **Rantala S, Vuopio-Varkila J, Vuento R, Huhtala H, Syrjänen J**. 2009. Clinical presentations and epidemiology of β-haemolytic streptococcal bacteraemia: a population-based study. Clin. Microbiol. Infect. **15**:286–288.

6. **Sylvetsky N, Raveh D, Schlesinger Y, Rudensky B, Yinnon AM**. 2002. Bacteremia due to beta-hemolytic *Streptococcus* group G: increasing incidence and clinical characteristics of patients. Am. J. Med. **112**:622–626.

7. **Takahashi T, Ubukata K, Watanabe H**. 2011. Invasive infection caused by *Streptococcus dysgalactiae* subsp. *equisimilis*: characteristics of strains and clinical features. J. Infect. Chemother. **17**:1–10.

8. **Brandt CM, Spellerberg B**. 2009. Human infections due to *Streptococcus dysgalactiae* subspecies *equisimilis*. Clin. Infect. Dis. **49**:766–772.

9. **Facklam R**. 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin. Microbiol. Rev. **15**:613–630.

10. **Whatmore AM, Engler KH, Gudmundsdottir G, Efstratiou A**. 2001. Identification of isolates of *Streptococcus canis* infecting humans. J. Clin. Microbiol. **39**:4196–4199.

11. **Ahmad Y, Gertz RE, Li Z, Sakota V, Broyles LN, Van Beneden C, Facklam R, Shewmaker PL, Reingold A, Farley MM, Beall BW**. 2009. Genetic relationships deduced from *emm* and multilocus sequence typing of invasive *Streptococcus dysgalactiae* subsp. *equisimilis* and *S. canis* recovered from isolates collected in the United States. J. Clin. Microbiol. **47**:2046–2054.

12. **Vandamme P, Pot B, Falsen E, Kersters K, Devriese LA**. 1996. Taxonomic study of lancefield streptococcal groups C, G, and L (*Streptococcus dysgalactiae*) and proposal of *S. dysgalactiae* subsp. *equisimilis* subsp. nov. Int. J. Syst. Bacteriol. **46**:774–781.

13. **Cohen-Poradosu R, Jaffe J, Lavi D, Grisariu-Greenzaid S, Nir-Paz R, Valinsky L, Dan-Goor M, Block C, Beall B, Moses AE**. 2004. Group G streptococcal bacteremia in Jerusalem. Emerging Infect. Dis. **10**:1455–1460.

14. **Tseng S-P, Lin Y-Y, Tsai J-C, Hsueh P-R, Chen H-J, Hung W-C, Teng L-J**. 2010. Distribution of *emm* types and genetic characterization of the *mgc* locus in group G *Streptococcus dysgalactiae* subsp. *equisimilis* from a hospital in northern Taiwan. J. Clin. Microbiol. **48**:2975–2977.

15. **Sunaoshi K, Murayama SY, Adachi K, Yagoshi M, Okuzumi K, Chiba N, Morozumi M, Ubukata K**. 2010. Molecular *emm* genotyping and antibiotic susceptibility of *Streptococcus dysgalactiae* subsp. *equisimilis* isolated from invasive and non-invasive infections. J. Med. Microbiol. **59**:82–88.

16. **Jensen A, Kilian M**. 2012. Delineation of *Streptococcus dysgalactiae*, its subspecies, and its clinical and phylogenetic relationship to *Streptococcus pyogenes*. J. Clin. Microbiol. **50**:113–126.

17. **Pinho MD**. 2012. Clonal diversity and antibiotic susceptibility of invasive *Streptococcus dysgalactiae* subsp. *equisimilis* isolated in Portugal. Oral Presentation O 553, London, United Kingdom.

18. **Kittang BR, Bruun T, Langeland N, Mylvaganam H, Glambek M, Skrede S**. 2011. Invasive group A, C and G streptococcal disease in western Norway: virulence gene profiles, clinical features and outcomes. Clin. Microbiol. Infect. **17**:358–364.

19. **Nei T, Akutsu K, Shima A, Tsuboi I, Suzuki H, Yamamoto T, Tanaka K, Shinoyama A, Kojima Y, Washio Y, Okawa S, Sonobe K, Norose Y, Saito R**. 2012. A case of streptococcal toxic shock syndrome due to Group G streptococci identified as *Streptococcus dysgalactiae* subsp. *equisimilis*. J. Infect. Chemother. **18**:919-924.

20. **Geyer A, Schmidt KH**. 2000. Genetic organisation of the M protein region in human isolates of group C and G streptococci: two types of multigene regulator-like (*mgrC*) regions. Mol. Gen. Genet. **262**:965–976.

21. **Minami M, Ichikawa M, Matsui H, Hata N, Wakiyama N, Matsumoto M, Ohta M, Hasegawa T**. 2011. Prevalence of a streptococcal inhibitor of a complement-mediated cell lysis-like gene (*sicG*) in *Streptococcus dysgalactiae* subsp. *equisimilis*. Curr. Microbiol. **62**:884–887.

22. **Shimomura Y, Okumura K, Murayama SY, Yagi J, Ubukata K, Kirikae T, Miyoshi-Akiyama T**. 2011. Complete genome sequencing and analysis of a Lancefield group G *Streptococcus dysgalactiae* subsp. *equisimilis* strain causing streptococcal toxic shock syndrome (STSS). BMC Genomics **12**:17.

23. **Richards VP, Zadoks RN, Bitar PD, Lefébure T, Lang P, Werner B, Tikofsky L, Moroni P, Stanhope MJ**. 2012. Genome characterization and population genetic structure of the zoonotic pathogen, *Streptococcus canis*. BMC Microbiol. **12**:293.

24. **McMillan DJ, Kaul SY, Bramhachari PV, Smeesters PR, Vu T, Karmarkar MG, Shaila MS, Sriprakash KS**. 2011. Recombination drives genetic diversification of *Streptococcus dysgalactiae* subspecies *equisimilis* in a region of streptococcal endemicity. PLoS ONE **6**:e21346.

25. **Yin J, Yu S, Liu X, Li Y, Gao W, Ma X, Yang Y**. 2011. Prevalence and mechanism of resistance to antimicrobial agents in group G streptococcal isolates from China. Antimicrob. Agents Chemother. **55**:402–404.

26. **Sriprakash KS, Hartas J**. 1996. Lateral genetic transfers between group A and G streptococci for M-like genes are ongoing. Microb. Pathog. **20**:275–285.

27. **Towers RJ, Gal D, McMillan D, Sriprakash KS, Currie BJ, Walker MJ, Chhatwal GS, Fagan PK**. 2004. Fibronectin-binding protein gene recombination and horizontal transfer between group A and G streptococci. J. Clin. Microbiol. **42**:5357–5361.

28. **Kalia A, Bessen DE**. 2004. Natural selection and evolution of streptococcal virulence genes involved in tissue-specific adaptations. J. Bacteriol. **186**:110–121.

29. **Choi SC, Rasmussen MD, Hubisz MJ, Gronau I, Stanhope MJ, Siepel A**. 2012. Replacing and additive horizontal gene transfer in *Streptococcus*. Mol. Biol. Evol. **29**:3309-3320.

30. **Simpson WJ, Musser JM, Cleary PP**. 1992. Evidence consistent with horizontal transfer of the gene (*emm12*) encoding serotype M12 protein between group A and group G pathogenic streptococci. Infect. Immun. **60**:1890–1893.

31. **Friães A, Ramirez M, Melo-Cristino J**. 2007. Nonoutbreak surveillance of group A streptococci causing invasive disease in Portugal identified internationally disseminated clones among members of a genetically heterogeneous population. J. Clin. Microbiol. **45**:2044–2047.

32. **Bessen DE, Carapetis JR, Beall B, Katz R, Hibble M, Currie BJ, Collingridge T, Izzo MW, Scaramuzzino DA, Sriprakash KS**. 2000. Contrasting molecular epidemiology of group A streptococci causing tropical and nontropical infections of the skin and throat. J. Infect. Dis. **182**:1109–1116.

33. **Lopardo HA, Vidal P, Sparo M, Jeric P, Centron D, Facklam RR, Paganini H, Pagniez NG, Lovgren M, Beall B**. 2005. Six-month multicenter study on invasive infections due to *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis* in Argentina. J. Clin. Microbiol. **43**:802–807.

34. **Kittang BR, Skrede S, Langeland N, Haanshuus CG, Mylvaganam H**. 2011. *emm* gene diversity, superantigen gene profiles and presence of SlaA among clinical isolates of group A, C and G streptococci from western Norway. Eur. J. Clin. Microbiol. Infect. Dis. **30**:423–433.

35. **Panchaud A, Guy L, Collyn F, Haenni M, Nakata M, Podbielski A, Moreillon P, Roten C-AH**. 2009. M-protein and other intrinsic virulence factors of *Streptococcus pyogenes* are encoded on an ancient pathogenicity island. BMC Genomics **10**:198.

36. **Suzuki H, Lefébure T, Hubisz MJ, Pavinski Bitar P, Lang P, Siepel A, Stanhope MJ**. 2011. Comparative genomic analysis of the *Streptococcus dysgalactiae* species group: gene content, molecular adaptation, and promoter evolution. Genome Biol. Evol. **3**:168–185.

37. **Davies MR, Tran TN, McMillan DJ, Gardiner DL, Currie BJ, Sriprakash KS**. 2005. Inter-species genetic movement may blur the epidemiology of streptococcal diseases in endemic regions. Microbes Infect. **7**:1128–1138.

38. **Sitkiewicz I, Green NM, Guo N, Mereghetti L, Musser JM**. 2011. Lateral gene transfer of streptococcal ICE element RD2 (region of difference 2) encoding secreted proteins. BMC Microbiol. **11**:65.

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

40. **Rato MG, Bexiga R, Nunes SF, Vilela CL, Santos-Sanches I**. 2010. Human group A streptococci virulence genes in bovine group C streptococci. Emerging Infect. Dis. **16**:116–119.

41. **Alber J, El-Sayed A, Estoepangestie S, Lämmler C, Zschöck M**. 2005. Dissemination of the superantigen encoding genes *seeL*, *seeM*, *szeL* and *szeM* in *Streptococcus equi* subsp. *equi* and *Streptococcus equi* subsp. *zooepidemicus*. Vet. Microbiol. **109**:135–141.

42. **Colón AE, Cole RM, Leonard CG**. 1971. Lysis and lysogenization of groups A, C, and G streptococci by a transducing bacteriophage induced from a group G *Streptococcus*. J. Virol. **8**:103–110.

43. **Colón AE, Cole RM, Leonard CG**. 1972. Intergroup lysis and transduction by streptococcal bacteriophages. J. Virol. **9**:551–553.

44. **Davies MR, McMillan DJ, Van Domselaar GH, Jones MK, Sriprakash KS**. 2007. Phage 3396 from a *Streptococcus dysgalactiae* subsp. *equisimilis* pathovar may have its origins in *Streptococcus pyogenes*. J. Bacteriol. **189**:2646–2652.

45. **Holden MTG, Heather Z, Paillot R, Steward KF, Webb K, Ainslie F, Jourdan T, Bason NC, Holroyd NE, Mungall K, Quail MA, Sanders M, Simmonds M, Willey D, Brooks K, Aanensen DM, Spratt BG, Jolley KA, Maiden MCJ, Kehoe M, Chanter N, Bentley SD, Robinson C, Maskell DJ, Parkhill J, Waller AS**. 2009. Genomic evidence for the evolution of *Streptococcus equi*: host restriction, increased virulence, and genetic exchange with human pathogens. PLoS Pathog. **5**:e1000346.

46. **Davies MR, Shera J, Van Domselaar GH, Sriprakash KS, McMillan DJ**. 2009. A novel integrative conjugative element mediates genetic transfer from group G streptococcus to other {beta}-hemolytic streptococci. J. Bacteriol. **191**:2257–2265.

47. **Figueiredo TA, Aguiar SI, Melo-Cristino J, Ramirez M**. 2006. DNA methylase activity as a marker for the presence of a family of phage-like elements conferring efflux-mediated macrolide resistance in streptococci. Antimicrob. Agents Chemother. **50**:3689–3694.

48. **Galimand M, Lambert T, Gerbaud G, Courvalin P**. 1999. High-level aminoglycoside resistance in the beta-hemolytic group G *Streptococcus* isolate BM2721. Antimicrob. Agents Chemother. **43**:3008–3010.

49. **Vojtek I, Pirzada ZA, Henriques-Normark B, Mastny M, Janapatla RP, Charpentier E**. 2008. Lysogenic transfer of group A *Streptococcus* superantigen gene among streptococci. J. Infect. Dis. **197**:225–234.

50. **Paillot R, Darby AC, Robinson C, Wright NL, Steward KF, Anderson E, Webb K, Holden MTG, Efstratiou A, Broughton K, Jolley KA, Priestnall SL, Marotti Campi MC, Hughes MA, Radford A, Erles K, Waller AS**. 2010. Identification of three novel

superantigen-encoding genes in *Streptococcus equi* subsp. *zooepidemicus*, *szeF*, *szeN*, and *szeP*. Infect. Immun. **78**:4817–4827.