Universidade de Aveiro Departamento de Biologia
2015

RICARDO JORGE DA SILVA CARVALHO

Reavaliação do diagnostico molecular de *Streptococcus pneumoniae*

Revisiting molecular diagnostics of *Streptococcus pneumoniae*

DECLARAÇÃO

Declaro que este relatório é integralmente da minha autoria, estando devidamente referenciadas as fontes e obras consultadas, bem como identificadas de modo claro as citações dessas obras. Não contém, por isso, qualquer tipo de plágio quer de textos publicados, qualquer que seja o meio dessa publicação, incluindo meios eletrónicos, quer de trabalhos académicos.

Universidade de Aveiro Departamento de Biologia
2015

RICARDO JORGE DA SILVA CARVALHO

Reavaliação do diagnostico molecular de *Streptococcus pneumoniae*

Revisiting molecular diagnostics of *Streptococcus pneumoniae*

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica da Doutora Raquel Sá-Leão, Investigadora principal no Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa e coorientação da Doutora Etelvina Figueira, Professora auxiliar do Departamento de Biologia da Universidade de Aveiro.

Dedico: á minha esposa: Ana; aos meus pais: Amadeu e Deolinda.

o júri

presidente

Prof. Doutora Maria de Lourdes Gomes Pereira
Professora associada com agregação da Universidade de Aveiro

Prof. Doutora Isabel da Silva Henriques
Investigadora auxiliar da Universidade de Aveiro

Prof. Doutora Raquel Sá-Leão Domingues da Silva
Investigadora principal da Universidade Nova de Lisboa

agradecimentos

Gostaria de agradecer, em primeiro lugar, à Doutora Raquel Sá-Leão pela oportunidade concedida de realizar o meu sonho de trabalhar e aprender Laboratory of Molecular Microbiology of Human Pathogens do Instituto de Tecnologia Química e Biológica da Universidade Nova de Lisboa. Para além de todo o apoio, paciência e compreensão que demonstrou durante todo o meu percurso. Obrigado por não ter desistido de mim! Agradeço ainda todas as oportunidades dadas durante todo o meu percurso. Deixo também o meu especial agradecimento à Débora Tavares que sempre me acompanhou e incutiu os melhores valores e técnicas de trabalho, estando presente para o que fosse preciso, e claro, obrigado pela generosidade e paciência. Não posso deixar de mencionar pessoas como: Sónia Almeida, Sofia Félix, Alexandra Simões, Carina Valente, Ana Cristina Paulo, Sara Handem que me acompanharam num dos momentos mais difíceis, dando todo o apoio possível, para que este trabalho fosse concluído, mesmo quando eu próprio duvidei. Um cumprimento especial para a Celine Coelho, também ela foi importante para eu me reorganizar dando-me força e incentivo.

O meu muito obrigado à Professora Hermínia de Lencastre e a todos os elementos do Laboratório de Genética Molecular.

O meu obrigado à Doutora Etelvina Figueira por tudo o que fez e me ajudou durante o meu percurso académico.

Agradeço aos meus pais por todo o sacrifício que fizeram dando tudo o que tinham, e muitas vezes o que não tinham, para que eu pudesse seguir e percorrer os caminhos por mim escolhidos apoiando-me incondicionalmente. Estou extremamente grato por tudo o que fizeram por mim.

O meu obrigado aos meus irmãos por todo o apoio dado e por me estenderem a mão quando eu mais precisava.

A toda a minha família, aquela que eu tinha antes de me casar e a que me acolheu como seu membro mesmo antes de me casar, o meu muito obrigado por todo o apoio.

Um cumprimento muito especial para o Sr. Augusto, que, sem a sua compreensão nada teria sido possível. Obrigado por toda a flexibilidade demonstrada e pela compreensão no atraso de algumas das minhas tarefas laborais. Muito obrigado!

Por fim, agradeço à mulher da minha vida, que estando a atravessar um momento único e de muita dificuldade teve forças para me dar a mão quando eu mais precisei. Superaste e ajudaste-me a superar!! És um exemplo de que nada é impossível!! Amo-te.

Muito Obrigado a todos os que me apoiaram!!!

palavras-chave

Streptococcus pneumoniae, identificação, *lytA*, *piaA*, *hylA*, qPCR, métodos moleculares

resumo

Streptococcus pneumoniae é uma bactéria patogénica que coloniza a nasofaringe humana. *S. pneumoniae* é responsável por causar doenças, tanto invasivas como não invasivas como: otite, pneumonia, meningite e sepsis, continuando a ser uma das principais causas de doenças infecciosas a nível mundial. Devido a semelhanças com espécies que lhe são estreitamente relacionadas, e que compartilham o mesmo nicho ecológico, pode ser um desafio identificar corretamente *S. pneumoniae* aplicando apenas técnicas não dependentes do passo de cultura bacteriana como a técnica de PCR em tempo real (qPCR). Em 2007, um método molecular para identificação de *S. pneumoniae* baseado num qPCR e tendo como alvo o gene da autolisina principal (*lytA*) de *S. pneumoniae* foi proposto por Carvalho e seus colaboradores. Desde então, este tem sido usado de uma forma sistemática por vários grupos. Em 2013, foi proposto por Trzcinszki e seus colaboradores o uso da lipoproteína ABC transportadora de ferro PiaA como alvo num qPCR. O *piaA* qPCR foi usado em paralelo com o *lytA* qPCR. Contudo, a presença de genes homólogos de *lytA* foi descrita em espécies filogeneticamente próximas, como *S. pseudopneumoniae* e *S. mitis,* e a presença do gene *piaA* não é ubíquo entre *S. pneumoniae*. O gene da proteína hyaluronato lyase (*hylA*) é descrito como sendo ubíquo a todas as estirpes de *S. Pneumoniae.* Este gene ainda não foi usado até ao momento como alvo para a identificação de *S. pneumoniae*. Assim o objectivo do nosso estudo foi a avaliação da especificidade, sensibilidade, valor positivo preditivo (VPP) e valor negativo preditivo (VNP) do método *lytA* e *piaA* qPCR; construção de *hylA* qPCR avaliando os mesmos parâmetros acima referidos; analise dos ensaios de uma forma independente e em conjunto, para poder retirar conclusões sobre qual o melhor gene alvo, ou alvos, a usar na identificação de *S. pneumoniae*. Foram testadas um total de 278 estirpes anteriormente caracterizadas: 61 *S. pseudopneumoniae*, 37 estirpes do grupo Viridans, 30 estirpes referência de outras espécies de *Streptococcus* e 150 estirpes de S. *pneumoniae*. A coleção usada incluía tanto estirpes obtidas em colonização como estirpes obtidas em doença. Através do método Multilocus sequence analysis (MLSA) verificámos que estirpes de *S. pseudopneumoniae* podem ser incorretamente identificadas como *S. pneumoniae* quando é utilizado o *lytA* qPCR. Ainda assim, os resultados mostraram que como alvo único, o gene *lytA* apresenta a melhor combinação de valores de especificidade, a sensibilidade, VPP e VNP sendo, respetivamente, 98.5%, 100.0%, 98.7% e 100.0%. A combinação de genes com a melhor combinação de valores de especificidade, sensibilidade, VPP e VNP foi *lytA* e *piaA*, com 100.0%, 93.3%, 97.9% e 92.6%, respetivamente. De realçar que pelo método MLSA verificámos que estirpes de *S. pseudopneumoniae* podem ser incorretamente identificadas como *S. pneumoniae* e algumas estirpes capsuladas (23F, 6B e 11A) e não-capsuladas de *S. pneumoniae* não são identificadas quando usada esta combinação de genes. O gene *hylA* como alvo único apresentou o valor de PPV mais baixo, todavia identificou corretamente todos os *S. pneumoniae*.

keywords

Streptococcus pneumoniae, identification, *lytA*, *piaA*, *hylA*, qPCR, molecular methods

abstract

Streptococcus pneumoniae is a human pathobiont that colonizes the nasopharynx. *S. pneumoniae* is responsible for causing non-invasive and invasive disease such as otitis, pneumonia, meningitis, and sepsis, being a leading cause of infectious diseases worldwide. Due to similarities with closely related species sharing the same niche, it may be a challenge to correctly distinguish *S. pneumoniae* from its relatives when using only non-culture based methods such as real time PCR (qPCR). In 2007, a molecular method targeting the major autolysin (*lytA*) of *S. pneumoniae* by a qPCR assay was proposed by Carvalho and collaborators to identify pneumococcus. Since then, this method has been widely used worldwide. In 2013, the gene encoding for the ABC iron transporter lipoprotein PiaA, was proposed by Trzcinzki and collaborators to be used in parallel with the *lytA* qPCR assay. However, the presence of *lytA* gene homologues has been described in closely related species such as *S. pseudopneumoniae* and *S. mitis* and the presence of *piaA* gene is not ubiquitous between *S. pneumoniae*. The hyaluronate lyase gene (*hylA*) has been described to be ubiquitous in *S. pneumoniae*. This gene has not been used so far as a target for the identification of *S. pneumoniae*. The aims of our study were to evaluate the specificity, sensitivity, positive predicted value (PPV) and negative predicted value (NPV) of the *lytA* and *piaA* qPCR methods; design and implement a new assay targeting the *hylA* gene and evaluate the same parameters above described; analyze the assays independently and the possible combinations to access what is the best approach using qPCR to identify *S. pneumoniae*. A total of 278 previously characterized strains were tested: 61 *S. pseudopneumoniae*, 37 Viridans group strains, 30 type strains from other streptococcal species and 150 *S. pneumoniae* strains. The collection included both carriage and disease isolates. By Mulilocus Sequence Analysis (MLSA) we confirmed that strains of *S. pseudopneumoniae* could be misidentified as *S. pneumoniae* when *lytA* qPCR assay is used. The results showed that as a single target, *lytA* had the best combination of specificity, sensitivity, PPV and NPV being, 98.5%, 100.0%, 98.7% and 100.0% respectively. The combination of targets with the best values of specificity, sensibility, PPV and NPV were *lytA* and *piaA*, with 100.0%, 93.3%, 97.9% and 92.6%, respectively. Nonetheless by MLSA we confirmed that strains of *S. pseudopneumoniae* could be misidentified as *S. pneumoniae* and some capsulated (23F, 6B and 11A) and non-capsulated *S. pneumoniae* were not Identified using this assay. The *hylA* gene as a single target had the lowest PPV. Nonetheless it was capable to correctly identify all *S. pneumoniae*.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY DATA

LIST OF ABREVIATIONS

°C	Celsius degree
µm	Micrometer
µM	Micro-molar
µl	Microliter
α-hemolysis	Alpha hemolysis
AW1	Wash solution 1
AW2	Wash solution 2
AE	Elution buffer
bp	Base par
C_T	Cycle threshold
CO_2	Carbon dioxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphatases
EDTA	Ethylenediamine tetraacetic acid
g	Grams
HIV	Human immunodeficiency virus
HylA	Hyaluronate lyase protein
hylA	Hyaluronate lyase gene
HA	Hyaluronic acid
HGT	Horizontal gene transfer
LytA	Autolysin protein
lytA	Autolysin gene
map	Methionine aminopeptidase
min	Minutes
ml	Milliliter
MLSA	Multilocus Sequence Analysis

mM	Milimolar
MgCl2	Magnesium chloride
ng	Nanograms
NPV	Negative predicted value
PCR	Polymerase chain reaction
pfl	Pyruvate formate lyase gene
pg	Pico grams
pH	Hydrogen potential
PiaA	ABC iron transporter lipoprotein
piaA	ABC iron transporter lipoprotein gene
Ply	Pneumolysin
ply	Pneumolysin gene
PsaA	Pneumococcal surface adhesin protein
psaA	Pneumococcal surface adhesin gene
ppaC	Inorganic pyrophosphatase gene
pyk	Pyruvate kinase gene
qPCR	Real-time polymerase chain reaction
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RPM	Rotation per minute
rpoB	RNA polymerase beta subunit gene
sec	Seconds
sodA	Superoxide dismutase gene
Tris.Cl	Trisaminometane chloride
TSA	Trypticase soy agar
tuf	Elongation factor Tu gene

1. INTRODUCTION

1.1 *Streptococcus pneumoniae* characterization

S. pneumoniae is a capsulated Gram-positive bacterium that can form diplococcus or shorter or longer chains of lancet-shaped cocci (Figure 1). The size of the chains is variable between 0.5 and 2µm. This bacterium is characterized by lacking the enzyme catalase, which cleaves hydrogen peroxide, and so is classified as catalase negative. This bacterium is also facultative anaerobic (1).

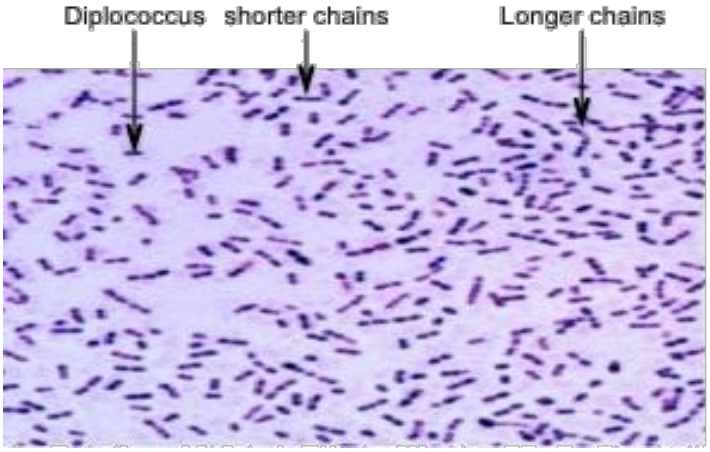

Figure 1. ***S. pneumoniae*** **view by optical microscope with Gram-positive coloration.**

When grown in laboratory conditions, it is a demanding and hard to please bacterium. The medium of growth needs to be supplemented, normally with sheep or horse blood for solid media and serum for liquid media. The optimal pH level is between 7 and 8, the optimal temperature is 37 °C and the atmosphere should be enriched with 5% of CO_2 (1). This bacterium, when grown in a medium supplemented with blood, has the ability to partially degrade blood cells, generating α-hemolysis (2).

1.2 *S. pneumoniae* epidemiology

S. pneumoniae is described as a frequent but transient commensal of the upper respiratory tract of humans, most precisely of the nasopharynx and oropharynx (Figure 2). Although described as a commensal microorganism it is also responsible for causing non-invasive and invasive disease remaining a leading cause of infections worldwide (3, 4). Carriage of *S. pneumoniae* in the upper respiratory tract is considered a prerequisite for the acquisition of pneumococcal disease (4). *S. pneumoniae* carriage is believed to be important for transmission of this pathobiont within the community, which explains why in crowding places, such as day-care centers and retirement homes the spread of pneumococcal strains increases (5, 6). Young children have the highest incidence of *S. pneumoniae* colonization; therefore it implies that this group is very important for dissemination of pneumococcal strains within the community (7).

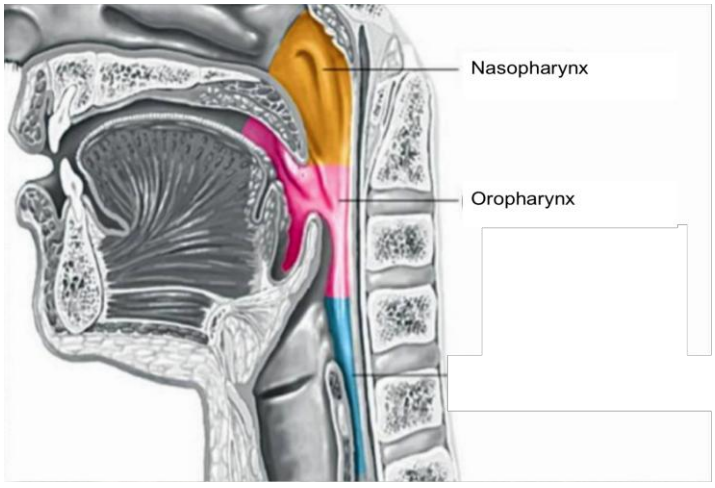

Figure 2. Representation of the ecological niches of *S. pneumoniae*: nasopharynx and oropharynx. (Adapted from Seeley and collaborators (8))

The rates of pneumococcal disease are disproportionally distributed among the human cycle of life. The two age groups with higher rates of pneumococcal disease are located at extremes: the young children (less than five years of age) and the elderly (over 60 years of age) (9, 10). However, apart from age there are other risk factors that can lead to *S. pneumoniae* infection: immunocompromised

patients, those with chronic diseases, HIV infection, cochlear implants, cerebrospinal fluid leak or smoking (11, 12).

In 2013 *S. pneumoniae* was responsible for 14.9% of the deaths that occurred in children below five years of age worldwide. Most of these cases were children who lived in developing countries and were below the age of 1 year (13).

Currently the strategy to prevent pneumococcal disease focus in the vaccination of the risk groups (10). To access the impact of vaccination and to apply the correct treatment in case of disease, the correct identification of *S. pneumoniae* is needed. The gold standard methods for the identification of live *S. pneumoniae* are conventional culture methods applied to a nasopharyngeal swab, which are the bile solubility and optochin test (14). However, non-culture dependent methods have improved sensitivity on the detection of *S. pneumoniae* (15, 16). Some of these methods are based in the identification of virulence factors, or on characteristics of these virulence factors such as particular DNA sequences (17-19).

1.3 *S. pneumoniae* virulence factors

An important virulence factor of *S. pneumoniae* is the polysaccharide capsule. There are more than 95 different serotypes with unique capsular polysaccharide structures (20). In fact, the capsule is extremely important for the survival of *S. pneumoniae* in the bloodstream (21-24).

A pneumococcal cell wall underlies the capsule, which consists of three major components, the lipoteichoic acid, the teichoic acid and the peptidoglycan. The teichoic acid has an important role in anchoring some of the surface proteins, which have a crucial role in colonization and inflammation (25). Between them are pneumolysin (Ply), autolysin (LytA), pneumococcal surface adhesin A (PsaA), Hyaluronate lyase (HylA) and ABC transporter lipoprotein (PiaA) that have been described as virulence factors of *S. pneumoniae*, having an important role in

colonization and in causing disease (26-31). Therefore, some of the genes encoding these proteins have been used to detect *S. pneumoniae* (18, 19).

A description of virulence factors of interest for the present work will follow.

1.3.1 LytA

LytA is the major autolysin of *S. pneumoniae* and is considered a major virulence factor of *S. pneumoniae* (27). It is a N-acetylmuramoyl-L-alanine amidase, responsible for the lytic activity of the *S. pneumoniae* (32). The *in vivo* function of this protein can be quite vast. Studies have shown that this protein presents an important role in releasing other virulence factors, such as Ply, being involved in *S. pneumoniae* pathogenesis. It can also release other proteins that will interact with the host and interfere in immune response (25, 33, 34). Another function of this protein is related with the ability of lysing other non-competent *S. pneumoniae*. This leads to a higher genetic diversity by increase the number of the competent *S. pneumoniae*, and therefore, increase the probability of exchanging DNA between *S. pneumoniae* and the assimilation of exogenous DNA (35).

Several studies have shown that attenuation of the LytA activity causes a decrease in virulence and, therefore a decrease in disease, caused by *S. pneumoniae* (27, 34).

This protein is encoded by the gene *lytA* that has been described to be present in the core genome of *S. pneumoniae*, being used for the identification of *S. pneumoniae* (36). However a few studies suggest that a homologous of this gene can be found in other Streptococcal species (36-38).

1.3.2 PiaA

PiaA is a lipoprotein that makes up one part of the two components of the *S. pneumoniae* iron uptake transport system and it is thought to be present on the

bacterium cell surface (39, 40). The role of this protein is related to *S. pneumoniae* iron acquisition, being essential to the full virulence of *S. pneumoniae*. However, the mechanism by which *S. pneumoniae* acquires iron is not fully understood. Brown and collaborators have shown the importance of this protein for the full virulence of *S. pneumoniae* (28).

PiaA is encoded by *piaA* gene. This gene has been described to be absent in other *Streptococcus* species and present only in *S. pneumoniae*. However Whalan and collaborators have shown that some non-capsulated *S. pneumoniae* strains lack the *piaA* gene (41).

1.3.3 HylA

Boulonois and collaborators have shown that HylA, encoded by the *hylA* gene, is present in the majority of pneumococcal isolates recovered from clinical samples (42). The function of this protein is promoting the degradation of the hyaluronic acid (HA), which is a major component of the human connective tissue (43, 44). The degradation of the HA brings advantages to the colonization and the subsequent infection, promoting the virulence of *S. pneumoniae* (27, 29).

However, studies performed with other bacterial species have shown that a functional HylA is not essential to cause disease (45, 46).

1.4 *S. pneumoniae* closely related species

The differentiation between *S. pneumoniae*, *S. pseudopneumoniae*, *S. mitis* and other closed related species can be difficult, since they are clustered in the same group, the Mitis group of Viridans Streptococci. The bacteria of this group are very promiscuous and the DNA can be shared between them (47, 48). A correct identification of this bacterium is extremely important in clinical diagnosis and in surveillance studies. However, due to the possibility of genetic exchanges

between these species, correct identification remains a challenge to microbiologists and taxonomists (2).

Closely related species of *S. pneumoniae* and the methodologies for the identification of *S. pneumoniae* relevant to this study will be described in the next sections.

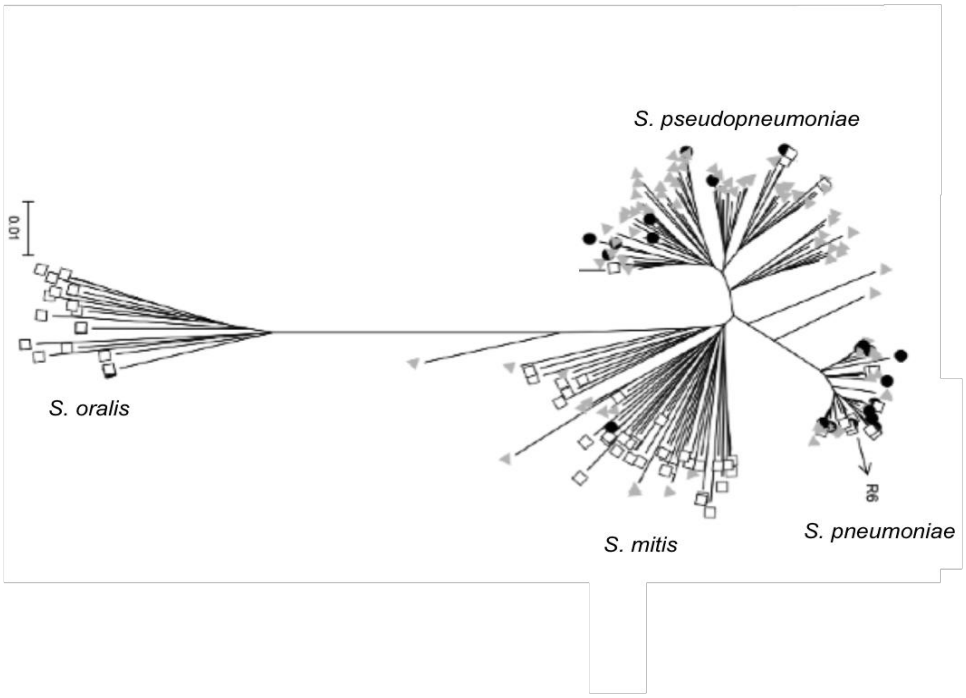

Figure 3. Multilocus Sequence Analysis (MLSA) phylogenetic based tree. (Adapted from Rolo and collaborators (49))

This MLSA phylogenetic based tree shows how *S. pneumoniae*, *S. pseudopneumoniae* and *S. mitis* are so closely related. The analysis was based on the concatenated sequences of six housekeeping genes: *aroE*, *gdh*, *gki*, *recP*, *spi* and *xpt*.

1.4.1 *S. pseudopneumoniae*

S. pseudopneumoniae is a recently described member of the Mitis group of Viridans Streptococci, which has some phenotypic and genetic characteristics identical to *S. pneumoniae*, *S. mitis* and *S. oralis*, but the average aminoacid

identity (AAI) between them is lower than 95%, and therefore has been considered a new species (50, 51).

S. pseudopneumoniae is catalase negative and with the shape of cocci. This bacterium differs from *S. pneumoniae* by being optochin resistant in the presence of 5% CO_2, and is bile insoluble. Another characteristic is the lack of a capsule, conferring a negative result for a biochemical reaction based on the use of a specific antibody targeting the capsule, named the Quellung reaction (50).

These strains are known to cause infections in patients who have a chronic obstructive pulmonary disease or exacerbation of chronic obstructive pulmonary disease (52).

1.4.2 *S. mitis*

S. mitis belongs to the Mitis group of Viridans streptococci, is catalase negative and with the shape of cocci. The cocci are arranged in chains. This bacterium differs from *S. pneumoniae* by being resistant to optochin, and like *S. pseudopneumoniae*, bile insoluble and have a negative result when submitted to the Quellung reaction (53). This bacterium is responsible for infections such as bacterial endocarditis (54).

1.4.3 Other Viridans group isolates

"*S. pneumoniae* – like" isolates belong to the Viridans group. These are isolates that are identified as putative *S. pneumoniae*. However, the result of at least one of the standard methods commonly used (optochin sensibility, bile solubility and Quellung reaction) in the identification of *S. pneumoniae* is divergent in these strains (17, 36). Sometimes species assignment of this isolates is extremely difficult (51, 55).

The origin of these bacteria has been studied and several hypotheses have been proposed. The members of the Viridans group are naturally competent for genetic transformation, and for this reason, they can incorporate exogenous DNA into their genome. Some studies refer to the fact that strains evolved from a common pathogenic ancestral by reduction of the genome, losing some virulence factors (56). This hypothesis is supported by the fact that *S. mitis*, *S. oralis* and *S. pneumoniae* share virulence factors (37).

Another hypothesis is related to Horizontal Gene Transfer (HGT) between species that share the same niche, which attenuates the difference between the species (50). In a recent study, Denapaite and collaborators have shown that HGT between *S. mitis* and *S. pneumoniae* occurs supporting this hypothesis (37).

1.5 *S. pneumoniae* identification methods

The traditional methods for *S. pneumoniae* identification are divided into two classes: the phenotypic-based methods and the genotypic-based methods.

Some of the methods that are important for the goals of this study will be described in the following sections.

1.5.1 Phenotypic-based methods

The phenotypic based methods are always culture dependent. This can lead to an under estimation of the rate of *S. pneumoniae* in the sample, due to the low sensitivity when compared with molecular methods (57).

Characteristics and limitations of the standard methods based on phenotypic characteristics in the identification of *S. pneumoniae* will be described next.

1.5.1.1 Optochin sensibility

Optochin is used in order to establish a differentiation between *S. pneumoniae* and other alpha-hemolytic streptococci. In this procedure optochin (ethylhydrocupreine hydrochloride) is used to test the fragility of the organism cell membrane (58).

In fact, *S. pneumoniae* is sensitive to this particular chemical compound when incubated in a 5% CO_2 atmosphere, while other alpha-hemolytic Viridans group bacteria like *S. mitis* and *S. pseudopneumoniae* are resistant to it. However, some pneumococci have been described as optochin resistant when incubated with the same conditions (36, 50, 55, 59-61).

S. pseudopneumoniae is characterized for being optochin resistant at 5% CO_2 and susceptible in normal atmosphere conditions (50). However more recently Rolo and collaborators have shown variations in this pattern (51).

These facts may create some difficulties in the distinction of *S. pneumoniae* from other closely related species, when only this method is applied.

1.5.1.2 Bile solubility

Bile solubility is considered by some researchers as one of the best methods to distinguish *S. pneumoniae* from the closely related species with a high sensitivity and a high specificity. *S. pneumoniae* is “soluble” (occurs the lyse o the cell) when submitted to a concentration of 2% of bile salts (62).

This method is based on the activity of the pneumococcal autolysins, such as LytA, promoted by bile salts (62). However, point mutations on the *lytA* gene can lead to a non-functional LytA protein. As a result, these strains have a different pattern, with no solubility when submitted to this phenotypic test (63). In addition, some members of Viridans group have been described as bile soluble (17).

1.5.1.3 The Quellung reaction

The Quellung reaction is the most common standard method for pneumococcal capsular serotyping, which involves testing a pneumococcal cell suspension in which specific antisera is used directly against the capsular polysaccharide (64). Nonetheless, not all pneumococcal strains are correctly identified by this method. Non-capsulated pneumococcal strains lack the capsule and therefore, this reaction will always be negative. Nonetheless, non-pneumococcal isolates with positive agglutination when polyvalent antiserum (omniserum) is used have been reported (65).

1.5.2 Genotypic-based methods

Several targets and methods were proposed to distinguish *S. pneumoniae* from the other closely related species. Genes encoding for LytA, PsaA and Ply described as being ubiquitous in *S. pneumoniae* were proposed for the identification and differentiation between *S. pneumoniae* and the related species (36, 66). However, homologous genes from the ones encoding these proteins were found in *S. pneumoniae* closely related species (38, 67).

Characteristics, advantages and limitations of standard molecular methods commonly used in routine laboratory identification of *S. pneumoniae* will be described next.

1.5.2.1 Multilocus Sequence Analysis for Viridans group (MLSA)

MultiLocus Sequence analysis has been used to establish relationships between species that are closely related or belong to the same genus. In some cases is used to infer if there is a new species, and therefore, a phylogenetic position of a new species can be established (68, 69).

Hoshino and collaborators using a concatenation of four housekeeping genes sequences (D-alanine:Dalanine ligase (*ddl*), glutamate dehydrogenase (*gdh*), RNA polymerase beta subunit (*rpoB*) and Superoxide dismutase (*sodA*)) first described the Viridans Group MLSA method. This method was developed to cluster and identify non-hemolytic streptococci (70). However, only the distant species were well resolved wile a poor resolution was obtained for close related species like *S. pseudopneumoniae* and *S. mitis* (69). Latter Bishop and collaborators developed a new scheme based on the concatenation of seven house-keeping sequences: Methionine aminopeptidase (*map*), Pyruvate formate lyase (*pfl*), Inorganic pyrophosphatase (*ppaC*), Pyruvate kinase (*pyk*), RNA polymerase beta subunit (*rpoB*), Superoxide dismutase (*sodA*) and elongation factor Tu (*tuf*). This scheme was able to well resolve all the species within the Viridans group. This method relies on comparison of the concatenated sequences in a web site (www.emlsa.net) and online database with concatenated sequences from species of Viridans group that were created for the propose (69).

There is enormous advantage by using this method to cluster strains, since it could be done by assessing an online database, and the strain is clustered almost immediately (69). Nonetheless, the web site is not always available and therefore this could be a disadvantage. Also only culturable strains can be submitted to this method witch results in a decrease of sensitivity (57). In addition is a laborious and expensive method for routine diagnosis (68).

1.5.2.2 Restriction Fragment Length Polymorphism (RFLP)

RFLP is a technique that allows the recognition of polymorphic sites in the DNA of a certain gene or organism by the use of restriction enzymes (71). These enzymes recognize a certain sequence and cut the sequence before, after or on the recognition site depending on the type of restriction enzyme used (72). The result is a pattern based on the size of the DNA fragments (71).

Lull and collaborators applied this method to distinguish *S. pneumoniae*, witch harbour a typical *lytA*, from the "*S. pneumoniae* - like" strains, that harbour

an atypical *lytA* (17). However, one *S. pseudopneumoniae* strain harbouring both typical and an atypical *lytA* has also been described, leading to misidentification between *S. pseudopneumoniae* and *S. pneumoniae* (38).

1.5.2.3 Real time PCR (qPCR)

The qPCR differs from the conventional PCR by the use of a labelled probe with a fluorophore, allowing the quantification of the target sequence and the analysis of the results without the need of a conventional electrophoresis (73). Another difference relies in the amount of DNA template necessary to amplify the product, which is considerably lower when compared to a conventional PCR (74).

Until today several target genes for identification of *S. pneumoniae* have been proposed. The gene *lytA* and a conjugation of *lytA* and *piaA* genes were employed, in order to avoid false positive results (18, 19).

When qPCR is used for organism identification, there is a need of a highly specific and sensitive probe to avoid the misidentification of bacterium (73). The lack of a highly specific and sensitive probe is the weakness of this assay.

2. AIMS

In recent years, identification of *S. pneumoniae* by qPCR using specific genes has been increasingly used in cultures and clinical samples (19, 75, 76). In 2007, Carvalho and collaborators described a qPCR targeting *lytA* gene (18). More recently, Trzcinszki and collaborators described a qPCR targeting the *piaA* gene to be used in parallel with the *lytA*, in order to increase the specificity of the assay (19).

The gene *hylA* has been described as ubiquitous in *S. pneumoniae* strains and absent in closely related species, such as *S. pseudopneumoniae* and *S. mitis*, nonetheless is was detected in one *S. oralis* strain (37, 38, 77). This gene has not been tested as a target to identify *S. pneumoniae*.

In this thesis we aimed to test and validate in an extensive collection of isolates the sensitivity, specificity, positive and negative predicted values of the qPCR assays targeting *lytA* and *piaA* genes; design and implement a qPCR targeting *hylA* gene, and test its sensitivity, specificity, positive and negative predicted values; analyze the assays independently and the possible combinations of the three targets to identify what is the best approach using qPCR method to identify *S. pneumoniae*.

3. METHODS

3.1 Study collection

A total of 278 isolates were used in this study (Table 1). To test the sensitivity and the specificity of the qPCR assay for genes *lytA*, *piaA*, and *hylA* a pneumococcal collection and a non-pneumococcal collection were used, respectively. As positive control TIGR4 pneumococcal strain was used and as a negative control the *S. pseudopneumoniae* strain ATCC BAA - 960 was used.

Table 1. Collection of isolates used in this study.

Species	Number isolates	Carriage/ disease	Year	Source	Country of origin
S. pneumoniae	150	Carriage	2001, 2006 and 2007	DCC	Portugal
S. pseudopneumoniae	61	Disease	1991 to 2009	Hospital	Spain
Streptococcus of Viridans group	37	Carriage and disease	1991 to 2009	DCC and hospital	Portugal and Spain
Type strains of *Streptococcus* species	30	NA	NA	NA	NA

DCC – Day care center; NA – Not available; Type strains of *Streptococcus* species: *S. gordonii* (DSM20568, DSM6777), *S. infantis* (DSM12492), *S. parasanguinis* (DSM6778), *S. peroris* (DSM 12493), *S. sanguinis* (DSM20567), *S. cinensis* (DSM14990), *S. vestibularis* (DSM5636), *S. oralis* (DSM20379, DSM20395), *S. cristatus* (DSM8249), *S. pseudopneumoniae* (ATCC BAA – 960, PT5479, IS7943), *S. equinus* (NCTC10389), *S. intermedius* (NCTC11324), *S. constellatus* (NCTC11325), *S. bovis* (DSM20480), *S. agalactiae* (DSMZ6784, DSM2134), *S. anginosus* (DSMZ20563), *S. canis* (DSM 20715), *S. dysgalactiae ssp. dysgalactiae* (DSMZ20662), *S. dysgalactiae ssp. equisimilis* (DSMZ6176), *S. equi ssp. zooepidemicus* (DSMZ20727), *S. mutans* (DSM20523), *S. pyogenes* (DSM 20565), *S. salivarius* (DSM20560), *Streptococcus spp* (*S. viridans* III and IV) and *S. pneumoniae* (TIGR4).

3.2 Growth conditions

A volume of 50 µl from each isolate, frozen in Mueller-Hinton broth with 15% of glycerol, was grown in Trypticase soy agar plates supplemented with 5% of

defibrinated sheep blood. Plates were incubated overnight at 37 °C with a 5% CO_2 atmosphere (78).

3.3 Genomic DNA extraction

Genomic DNA extraction was performed using the DNeasy Blood &Tissue Kit (Qiagen) according to a modified protocol: bacterium were collected with a loop and added to a microcentrifuge tube containing 172.8 µl of Tris.Cl pH 8 20 mM, sodium EDTA 2 mM, and 1.2% Triton X-100. After a pulsed vortex 7.2 µl of 1 mg/ml lysozyme was added to the previous solution and vortexed for 15 sec. Samples were incubated in a water bath for 60 min at 37 °C. A volume of 3 µl of RNAse 10 mg/ml was added to the samples and incubated for 5 min at room temperature. A volume of 25 µl of proteinase K was added to the solution and a series of pulsed vortexes was performed during 15 sec. Samples were incubated for 30 min at 56 °C. A volume of 200 µl of ethanol with a concentration of 99% stored at -20 °C was added to the solution and a series of pulsed vortexes was performed for 15 sec. The solution was transferred to a set comprising one spinning column and one collector tube and centrifuged at 8000 rpm for 1 min. The collector tube was discarded and the spinning column placed in a new collector tube. A volume of 500 µl of AW1 kit solution was added to the spinning column and incubated for 1 min at room temperature. The set was centrifuged at 8000 rpm for 1 min. The collector tube was discarded, and the same procedure was performed for AW2 kit solution with the exception of the centrifugation step, that was performed at 14000 rpm for 4 min. Next the collector tube was discarded and the DNA was eluted twice in 100 µl of the elution buffer (AE) to a 1.5 ml microcentrifuge tube. The first elution of DNA was stored at 4 °C and the second elution was stored at -20 °C. The concentration and quality of the DNA was measured using NanoDrop 1000 (Thermo scientific).

3.4 Real time PCR (qPCR)

Three qPCR assays were performed for the three genes under study: *lytA*, *piaA* and *hylA*. The primers and the probes used on the assays are in Supplementary Table 1 in the appendix.

Detection of *lytA* was performed as described by Carvalho and collaborators (18).

Detection of *piaA* was performed as described by Trzcinski (19) using primers and probe described by Walan and collaborators (41).

Detection of *hylA* gene was performed using primers and probe designed in this study. The primers were constructed based on the published sequence of the *hylA* gene (SP_0314) of *S. pneumoniae* TIGR4 strain (GenBank accession number: NC_003028.3). This sequence was considered the reference sequence. The reference sequence of *hylA* gene was blasted against the NCBI database. TIB MOLBIOL Syntheselabor GmbH constructed the primers and probe. To label the probe we chose Cy5.

The qPCR assays were carried out with a final reaction volume of 25 µl containing FastStart universal Probe Master, primer forward, primer reverse, probe, miliQ water and 2.5 µl of DNA sample per reaction. The final concentration of the primers was 0.15 µM and of the probe 0.075 µM. A non-template control (blank), a positive control (*S. pneumoniae* TIGR4) and a negative control (*S. pseudopneumoniae* ATCC BAA - 960) were used in all assays. Preparation of all reagents, master mix and sample distribution was performed in a biological safety cabinet class II.

The amplification reaction was carried out in CFX96™ Real time System (Bio-Rad) using the following cycling parameters: 1 cycle at 50 °C for 2 min, 1 cycle at 95 °C for 10 min, 45 cycles at 95 °C for 15 sec plus annealing temperature of the primer (60 °C for *lytA* and *piaA* gene; 55 °C for *hylA* gene) for 1 min.

To determine if an isolate was positive or negative for the gene tested a cycle threshold was defined. All isolates with a C_T value below 35 were considered

positive. For C_T values equal or higher than 35 isolates were considered negative. All 98 non-pneumococcal isolates with a positive amplification were retested a second time in a different day. The 30 type strains of other *Streptoccocus* species and the pneumococcal isolates were tested twice in two different days. If the results were not concordant the isolates were retested a third time.

To determine the minimum DNA concentration to be used in the assays, the sensitivity of the qPCR was calculated based on Burn and collaborators and Kim and collaborators under the assumption that: (i) the size of the genome of the tested isolates was equal to the size of the R6 S. pneumoniae strain, i. e, 2,400,000 bp; 1 pg contains $0.97x10^9$ bp; (iii) at least 22 copies of the genome are needed to have DNA amplification (74, 79). According to these assumptions the minimum DNA concentration to be used needs to be higher than $5.4x10^{-5}$ ng/µl.

We set a maximum DNA concentration of 0.2 ng/µl for all isolates based on the CFX96™ Real time System manufacturer information in order to avoid false positives. A dilution was performed in order to obtain a concentration of 0.2 ng/µl for isolates with higher DNA concentration.

3.5 MLSA for Viridans group

The amplification by PCR of internal fragments of seven housekeeping genes (Methionine aminopeptidase (*map*), Pyruvate formate lyase (*pfl*), Inorganic pyrophosphatase (*ppaC*), Pyruvate kinase (*pyk*), RNA polymerase beta subunit (*rpoB*), Superoxide dismutase (*sodA*) and elongation factor Tu (*tuf*)) was performed as previously described using primers indicated in Supplementary Table 2 in the Appendix (69). The *sodA* reverse primer used was modified in this study.

PCR reactions for MLSA were carried out in a total volume of 50 µl, containing: GoTaq at a final concentration of 0.02 U/µ; buffer Go Taq at a final concentration of 1 x; $MgCl_2$ at a final concentration of 5 mM; dNTPs at a final concentration of 0.08 mM each; primer forward and reverse at a final

concentration of 0.4 µM each, and MiliQ water. The thermal cycling reaction conditions were an initial denaturation at 94 °C for 4 min; 30 cycles of denaturation at 95 °C for 30 sec, annealing temperature 50 °C for 30 sec, and extension at 72 °C for 45 sec; final extension at 72 °C for 10 min. The annealing temperature is indicated in Supplementary Table 2 in the Appendix (69).

Sequencing was performed at Macrogen, Inc**.** The sequences were compared with a template sequence of each gene downloaded from www.emlsa.net, using the DNAStar to ensure that they have the same length as the template (trimming).

A concatenated sequence with all of the seven sequences from the housekeeping genes from each isolate was performed using the program EditSeq from the DNAStar software.

To perform the MLSA the concatenated sequences were align by the ClustalW method with the concatenated sequences from other well resolved strains belonging to the Viridans group, that were downloaded from the website www.emlsa.net. To perform the alignment the software MEGA 6 was used. The alignment was save in a MEGA Format. The phylogenetic tree was constructed based on the performed alignment. The minimum evolution parameter was used in the construction of the phylogenetic tree.

4. RESULTS

To assess what would be the best method for the identification of *S. pneumoniae*, the specificity and the sensitivity of the qPCR assays targeting the *lytA*, *piaA* and *hylA* genes were assessed using a collection of pure cultures of *S. pneumoniae* (expected to have all these genes) and non-pneumococcal strains from closely related species (expected to lack all these genes). The specificity was tested using 128 isolates from closely related species such as *S. pseudopneumoniae*, Viridans group isolates and type strains of *Streptococcus* species. The sensitivity was tested using 150 *S. pneumoniae* strains of diverse serotypes and genotypes (Table 1 and Supplementary Table 4). When unexpected results were obtained, confirmatory tests were performed as described in Figure 4.

4.1 Specificity and sensitivity of *lytA* qPCR assay

Of the 128 non-pneumococcal isolates used to test the specificity of the *lytA* qPCR assay, 126 were negative and two were positive (Figure 5A) with a C_T average of 23 and 33. The positive isolates were 8615 and 9111, respectively (Table 2 and Figure 5A). The positive isolates were identified as *S. pseudopneumoniae* by Rolo and collaborators (Table 2) (51). Using this collection, the specificity of the *lytA* qPCR assay was 98.5%.

A confirmation at species level of the positive strains 8615 and 9111 was performed using MLSA scheme. The result was concordant; the 8615 and 9111 isolates were identified as *S. pseudopneumoniae* (Table 2 and Figure 6).

Concerning the sensitivity of the *lytA* qPCR assay, all 150 pneumococcal isolates tested were positive as expected (Figure 5B). Using this collection, the sensitivity of the *lytA* qPCR assay was therefore 100.0%.

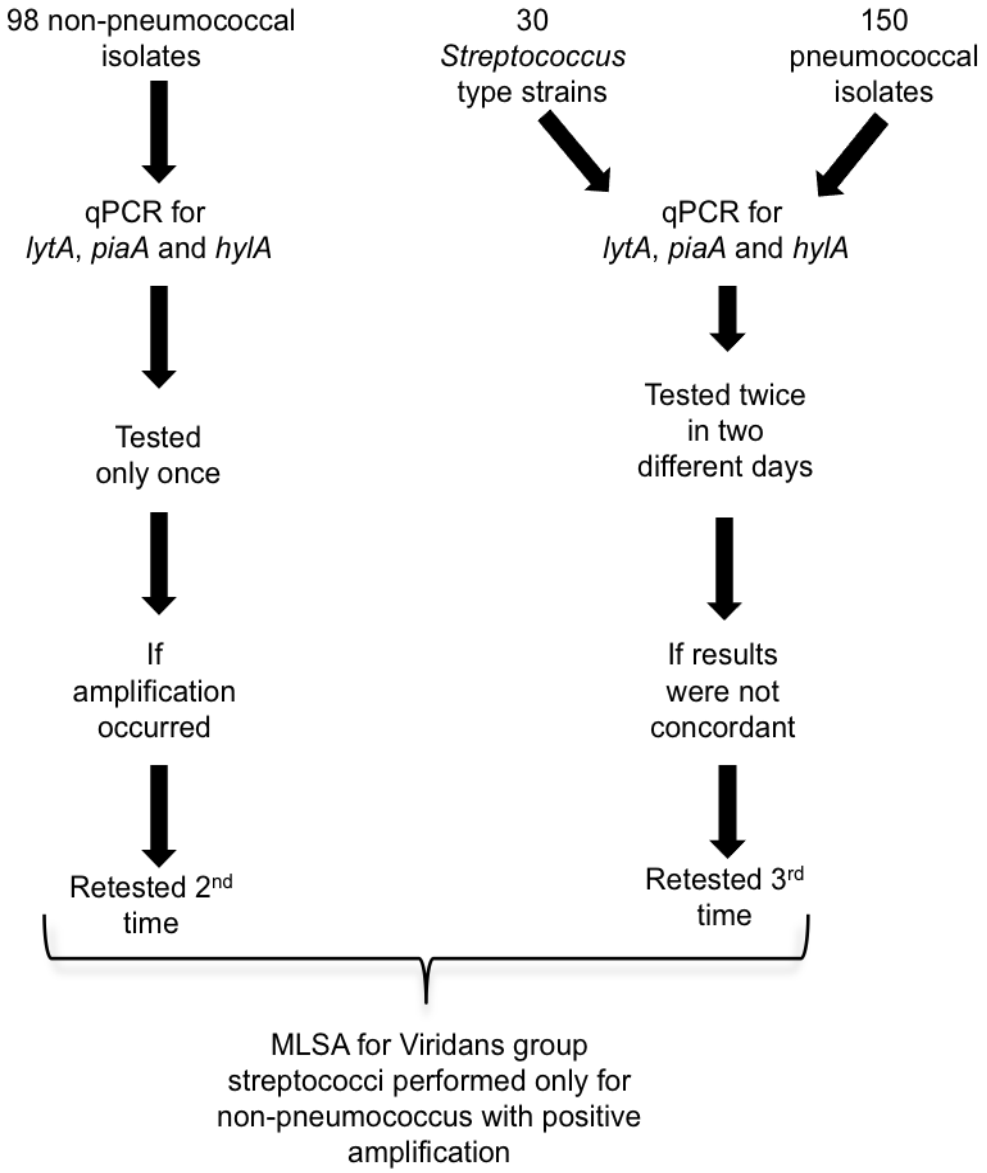

Figure 4. Flowshart of the methodology used during the study.

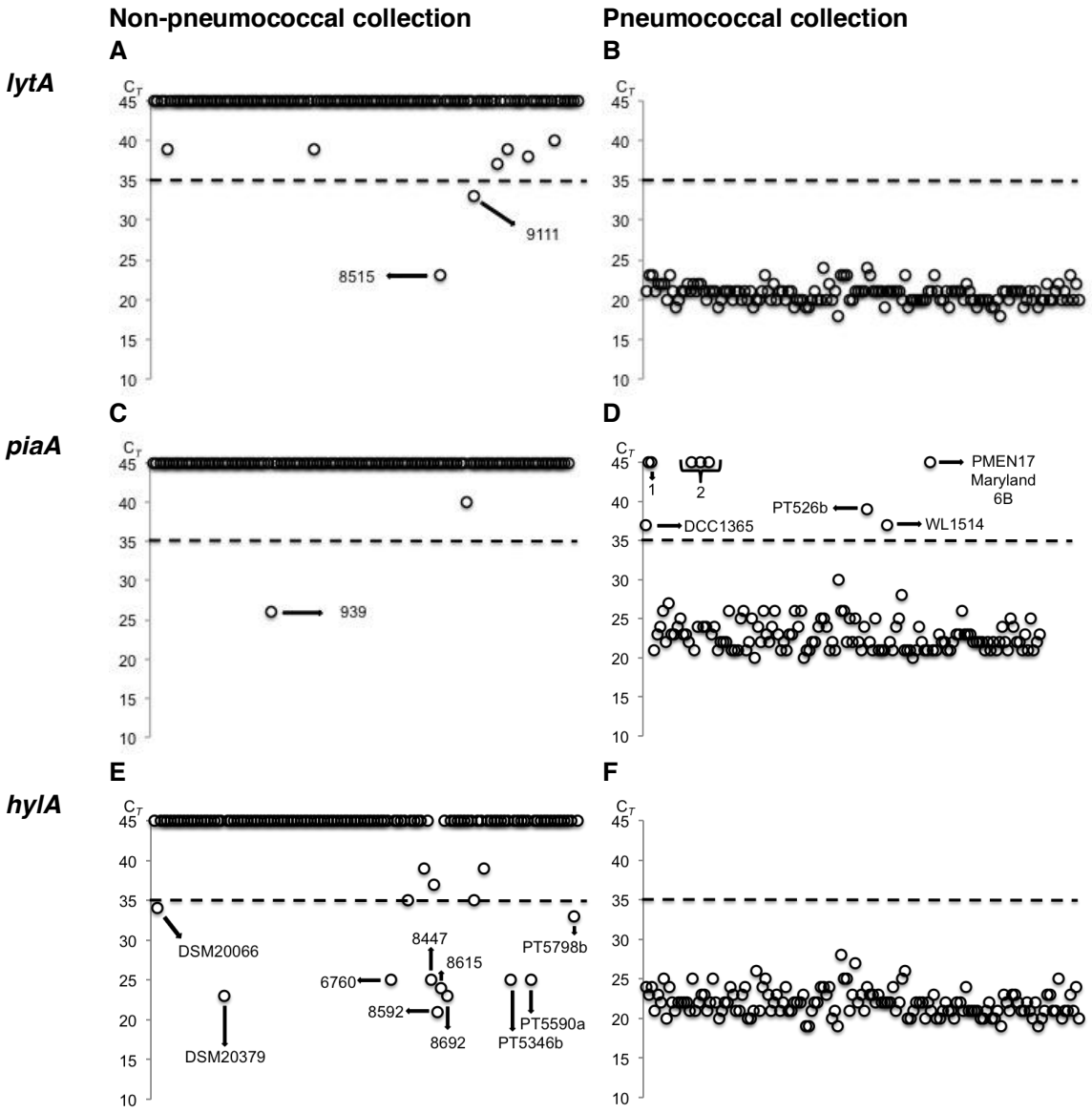

Figure 5. Specificity and sensitivity of the *lytA*, *piaA*, and *hylA* qPCR assays.

C_T value for the non-pneumococcal (A, C, and E) and pneumococcal (B, D and F) isolates for genes *lytA* (A and B) *piaA* (C and D) and *hylA* (E and F) are represented. The doted line represents the defined C_T threshold (C_T = 35) used to classify samples as positive or negative for a given assay. All isolates below the black doted line are positive and all isolates above the grey line are negative. With the number 1 are represented the isolate DCC1365, DCC635 and DCC646. The number 2 represents isolates PT1718, PT1804b and PT2293b.

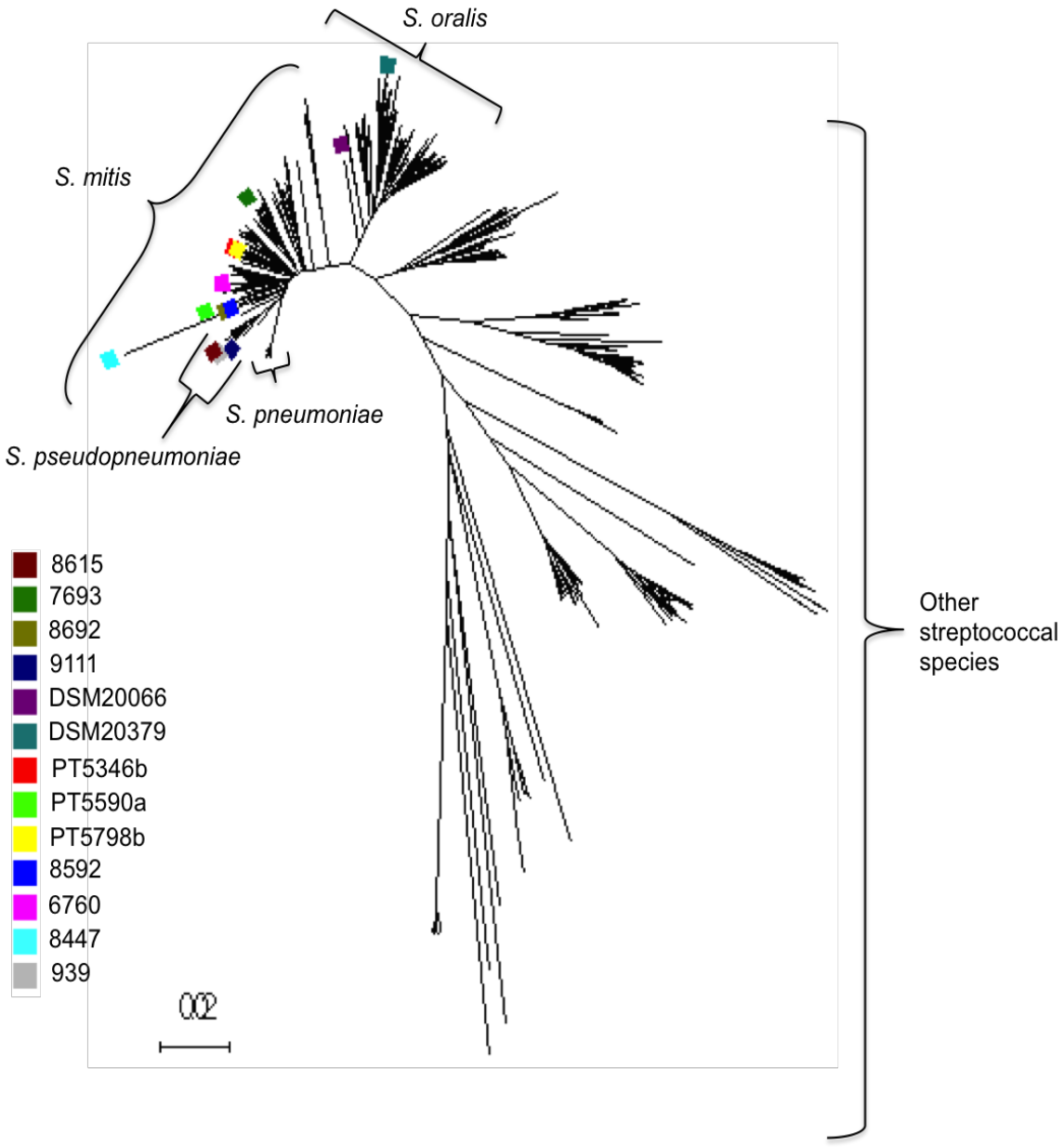

Figure 6. Phylogenetic position of the strains with atypical result based on a Multilocus Sequence Analysis for Viridans group.

In this figure a phylogenetic tree of the species of Viridans group is represented. This tree was performed by using the MLSA Viridans group scheme and a database of isolates available on www.eMLSA.net. The isolates with atypical results on the qPCR are represented by the coloured squares. With this analysis we were able to obtain a correct identification of all isolates tested at the species level in order to discuss the atypical results.

Table 2. Characterization of the strains with atypical results in this study.

Strain	Species	Carriage/ disease	Child/ adult	qPCR results	MLSA Result
8615	*S. pseudopneumoniae*	Disease	Adult	*lytA*$^{+}$ *hylA*$^{+}$	*S. pseudopneumoniae*
9111	*S. pseudopneumoniae*	Disease	Adult	*lytA*$^{+}$	*S. pseudopneumoniae*
939	*S. pseudopneumoniae*	Disease	Adult	*piaA*$^{+}$	*S. pseudopneumoniae*
6760	Viridans group	Disease	Adult	*hylA*$^{+}$	*S. mitis*
7693	Viridans group	Disease	Adult	*hylA*$^{+}$	*S. mitis*
8447	Viridans group	Disease	Adult	*hylA*$^{+}$	*S. mitis*
8592	Viridans group	Disease	Adult	*hylA*$^{+}$	*S. mitis*
8692	Viridans group	Disease	Adult	*hylA*$^{+}$	*S. mitis*
PT5346b	Viridans group	Carriage	Child	*hylA*$^{+}$	*S. mitis*
PT5590a	Viridans group	Carriage	Child	*hylA*$^{+}$	*S. mitis*
PT5798b	Viridans group	Carriage	Child	*hylA*$^{+}$	*S. mitis*
DSM20066	*S. oralis*	–	–	*hylA*$^{+}$	*S. oralis*
DSM20379	*S. oralis*	–	–	*hylA*$^{+}$	*S. oralis*

Bold – invasive strain; $^{+}$ positive; $^{-}$ negative; - No data.

4.2 Specificity and sensitivity of *piaA* qPCR assay

Of the 128 non-pneumococcal isolates used to test the specificity of the *piaA* qPCR assay, 127 were negative and one isolate was positive, 939 (Table 2 and Figure 5C). The positive isolate was *S. pseudopneumoniae* (Table 2 and Supplementary table 4). However, this strain was not the same as the ones that were positive for *lytA* gene (Table 2). Using this collection, the specificity of the *piaA* qPCR assay was 99.2%.

A confirmation at species level of the positive strain was performed using the Viridans group MLSA scheme. The result was concordant and therefore, the isolate was identified as *S. pseudopneumoniae* (Table 2 and Figure 6).

Regarding the sensitivity of the *piaA* qPCR assay, 140 out of 150 pneumococcal isolates were positive and the remaining ten isolates were negative (Figure 5D). Among the ten negative isolates, six were non-capsulated pneumococcal strains, two were of capsular type 23F, one was of capsular type 11A and one was a PMEN strain with capsular type 6B. Overall, the sensitivity of the *piaA* qPCR assay was 93.3%.

4.3 *hylA* qPCR assay

4.3.1 Design of *hylA* primers and probe

S. pneumoniae TIGR4 strain was used as a reference for the design of the primers and probe to be used in a *hylA* qPCR assay (TIGR4 annotation for *hylA*: SP_0314). A BLAST analysis of the TIGR4 *hylA* sequence was performed against the NCBI database. A total of 62 matches were found: from these 23 had a sequence homology of 100% between them, and therefore, we only used one representative sequence. In total, 34 sequences belonging to *S. pneumoniae* strains with an identity of 100%, including TIGR4 *S. pneumoniae* strain, and six sequences belonging to other three *Streptococcus* species (two *S. intermedius* strains, three *S. constellatus* strains and one *S. anginosus* strain) with a sequence identity of less than 76% were used.

In order to analyse the differences between the sequences, a BLAST analysis was performed between the 33 pneumococcal sequences and the other six sequences belonging to the *Streptococcus* genera with the reference sequence from the TIGR4 strain. A region of the gene was found to be highly conserved among the pneumococcal strains. On the other hand, this region was poorly conserved among the other three species. This region was 146 bp long and was located closer to the 5' end of the gene (Figure 7).

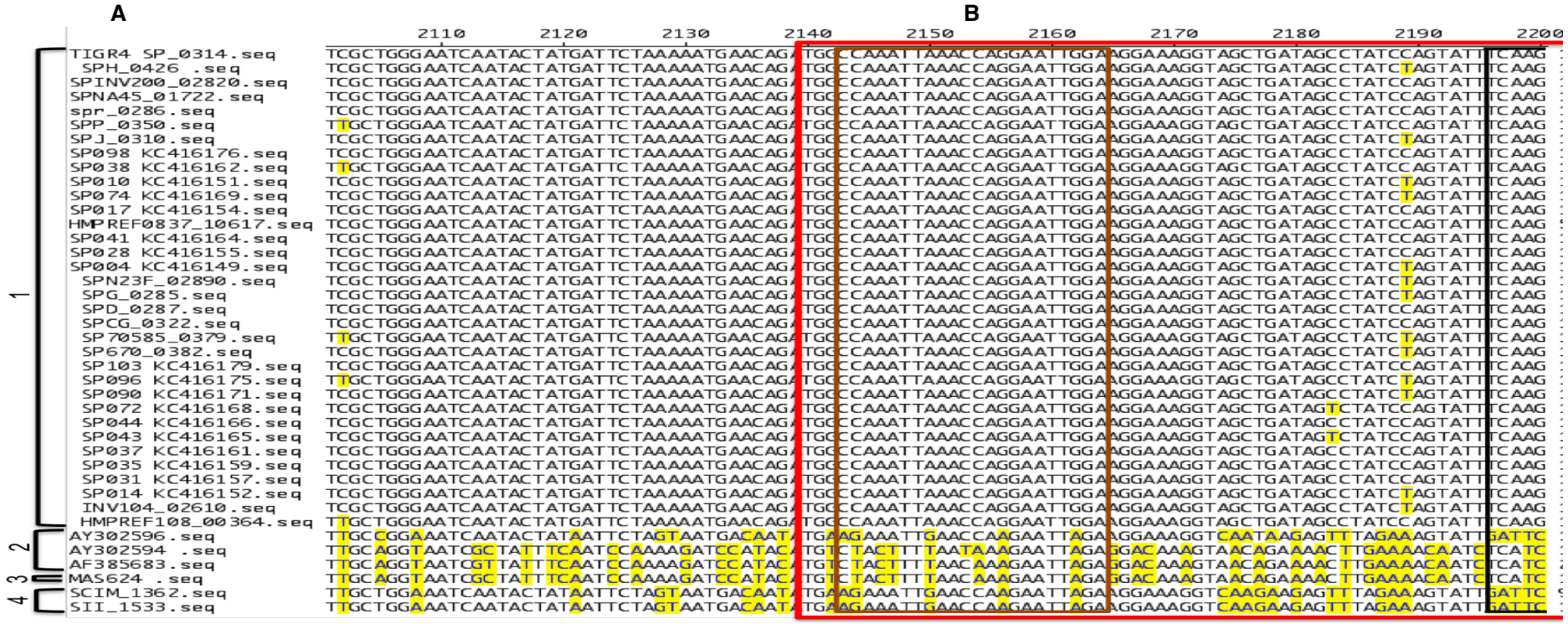

Figure 7. Sequence blast of the *hylA* gene sequence from TIGR4 *S. pneumoniae* strain (annotated as SP_0314 hyaluronate lyase) with other *hylA* gene sequence of 33 *S. pneumoniae* strains and six other *Streptococcus* species.

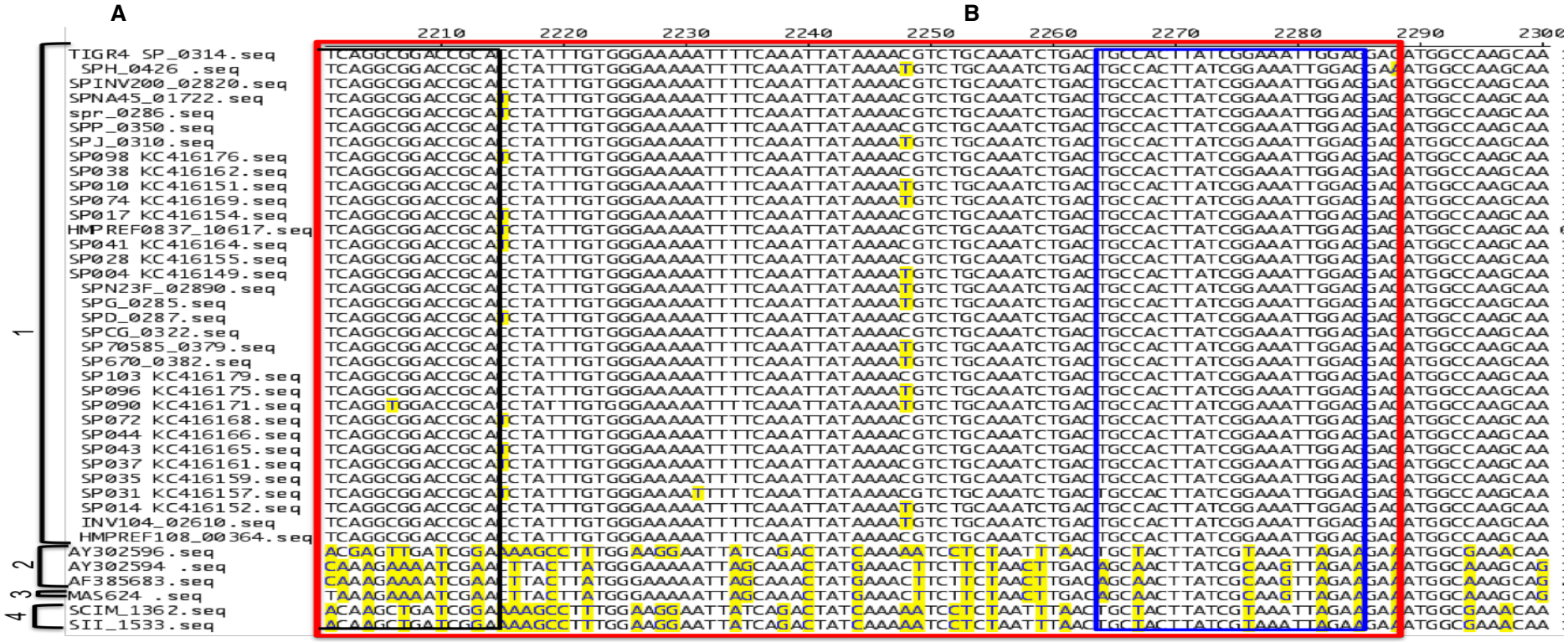

Figure 7. Sequence blast of the *hylA* gene sequence from TIGR4 *S. pneumoniae* strain (annotated as SP_0314 hyaluronate lyase) with other *hylA* gene sequence of 33 *S. pneumoniae* strains and six other *Streptococcus* species. (Cont.)

A – Locus Tag; B – Sequence of the strains blasted; 1- *S. pneumoniae* sequences; 2- *S. constellatus* sequences; 3- *S. anginosus* sequence; 4- *S. intermedius* sequences; Red rectangle is the chosen sequence as target for construction of primers and probe to be used in the qPCR; Yellow - non-consensus between the sequences; Brown box – primer forward; Black box – probe; blue box – primer reverse. The alignment was performed with MegAlign from DNAStar.

Having chosen the target sequence, two sets of primers and one probe were designed by TIB MOLBIOL Syntheselabor GmbH (Supplementary Table 3 in the appendix section). Both sets of primers were tested with two pneumococcal strains (TIGR4 and D39) as positive controls and one *S. mitis* strain (DSM12643) as a negative control. Three different annealing temperatures were tested: 60°C, 58°C and 55°C. The primers hylA-1F and hylA-1R were the best set of primers (Table 3), which yield a PCR product of 169 bp with the annealing temperature of 55°C.

Table 3. C_T values for *hylA* qPCR optimization for three different temperatures: 55°C, 58°C and 60°C.

Strain	Annealing temperature	Primers	C_T
D39	55 °C	hylA_1F + hylA_1R	13.9
		hylA_2F + hylA_2R	15.7
TIGR4	55 °C	hylA_1F + hylA_1R	22.9
		hylA_2F + hylA_2R	23.6
DSM12643	55 °C	hylA_1F + hylA_1R	<45
		hylA_2F + hylA_2R	40.2
D39	60 °C	hylA_1F + hylA_1R	25.1
		hylA_2F + hylA_2R	13.8
TIGR4	60 °C	hylA_1F + hylA_1R	32.1
		hylA_2F + hylA_2R	22.2
DSM12643	60 °C	hylA_1F + hylA_1R	<45
		hylA_2F + hylA_2R	37.2
D39	58 °C	hylA_1F + hylA_1R	16.1
		hylA_2F + hylA_2R	<45
TIGR4	58 °C	hylA_1F + hylA_1R	23.6
		hylA_2F + hylA_2R	22.1
DSM12643	58 °C	hylA_1F + hylA_1R	43.8
		hylA_2F + hylA_2R	36.2

The primers and temperatures were tested in two different *S. pneumoniae* strains, D39 and TIGR4 and in one *S. mitis* strain, DSM12643; hylA_1F + hylA_1R – first set of primers; hylA_2F + hylA_2R – second set of primers.

4.3.2 Specificity and sensitivity of *hylA* qPCR assay

Of the 128 non-pneumococcal isolates used to test the specificity of the *hylA* qPCR assay, 117 were negative and eleven were positive, DSM20066, DSM20379, 6760, 7693, 8447, 8592, 8615, 8692, PT5346b, PT5590a and PT5798b. (Table 2 and Figure 5E). From these positive isolates eight (6760, 7693, 8447, 8592, 8692, PT5346b, PT5590a and PT5798b) were from Viridans group isolates, two *S. oralis* strains (DSM20066 and DSM20379) and one was classified as *S. pseudopneumoniae* (8515) by Rolo and collaborators (51). Using this collection, the specificity of the *hylA* qPCR assay was 92.0%.

A confirmation at species level of the positives strains was performed using the MLSA scheme. The result was concordant and therefore, the eight isolates were identified as *S. mitis*, one as *S. pseudopneumoniae* and two as *S. oralis* (Table 2 and Figure 6).

Regarding the sensitivity of the *hylA* qPCR assay, 150 out of 150 pneumococcal isolates were positive therefore the sensitivity of the *hylA* qPCR assay was 100.0% (Figure 5F).

4.4 Combined specificity and sensitivity of the *lytA*, *piaA* and *hylA* qPCR assays

The combined analysis was performed in order to compare and access what would be the best target or targets to correctly identify *S. pneumoniae*. For a non-pneumococcal isolate to be considered positive (identified as *S. pneumoniae*) the qPCR result had to be positive for all the genes in the combination, on the other hand if one isolate was positive just for one of the genes the isolate was not considered to be *S. pneumoniae*. A pneumococcal isolate was considered negative (identified as non-*S. pneumoniae*) when the qPCR result was negative for at least one of the genes in the combination.

Table 4. Specificity and sensitivity for the genes tested.

Target genes	Specificity	Sensitivity
lytA	98.5% (126/128)	100.0% (150/150)
piaA	99.2% (127/128)	93.3% (140/150)
hylA	92.0% (117/128)	100.0% (150/150)
lytA piaA	100.0% (128/128)	93.3% (140/150)
lytA hylA	99.2% (127/128)	100.0% (150/150)
piaA hylA	100.0% (128/128)	93.3% (140/150)
lytA piaA hylA	100.0% (128/128)	93.3% (140/150)

Despite the fact that two non-pneumococcal isolates were positive for the *lytA* gene and one was positive for the *piaA* gene (Table 4 and Table 5) none of these isolates were positive for both genes. Therefore, the specificity of the combined assay was 100.0% (Table 4). Despite the fact that all isolates were correctly identified in the *lytA* qPCR assay, in the *piaA* qPCR assay ten pneumococci (6 non-typeable strains, two 23F strain, one 6B strain and one 11A) isolates were negative (Table 5). Therefore the sensitivity for the combined assay was 93.3% (Table 4)

Table 5. Results for the combination of the three qPCR assays targeting the genes *lytA*, *piaA* and *hylA*.

	S. pneumoniae	*S. pseudopneumoniae*	Viridans group strains	Type strains of *Streptococcus* species (does not include *S. pneumoniae*)
$lytA^+ piaA^+ hylA^+$	140	0	0	0
$lytA^- piaA^- hylA^-$	0	58	29	28
$lytA^+ piaA^- hylA^-$	0	1	0	0
$lytA^+ piaA^- hylA^+$	10	1	0	0
$lytA^- piaA^+ hylA^-$	0	1	0	0
$lytA^- piaA^- hylA^+$	0	0	8	2
Total	150	61	37	30

The data in the table corresponds to absolute numbers.

When the *lytA* and *hylA* qPCR assays were combined, the specificity was 99.2% (Table 4). Between the positive non-pneumococcal isolates, one was common for both genes, 8615 being positive for the *lytA* and *hylA* genes. The sensitivity for this combined assay was 100.0% (Table 4). Here the *hylA lytA* qPCR were able to detect all of 150 pneumococcal isolates.

Regarding the combination between the *hylA* and *piaA* qPCR assays, the value of the specificity of these combined assays is 100.0% (Table 4). The sensitivity for this combination is 93.3% (Table 4). From the 150 isolates tested ten were negative for *piaA* gene and none for *hylA* gene (Table 5).

With the combination between the three assays *lytA*, *piaA* and *hylA* qPCR, 14 out of 128 isolates were positive for, at least, one of the three genes in combination. One isolate was positive for the *lytA* and *hylA* genes but not for the *piaA* gene. Therefore, the combined specificity of these three assays is 100.0% (Table 4). The sensitivity for these combined assays was 93.3% (Table 4). In this case, all isolates are positive for the *lytA* and *hylA* qPCR but ten isolates are negative for the *piaA* qPCR (Table 5).

Overall, with the analyses of these results the method that combines the best value of specificity and sensitivity is the combination between the *lytA* and *hylA* qPCR assay with a specificity of 99.2% and a sensitivity of 100.0% (Table 4). Nonetheless it is necessary to take in account the PPV and NPV. The combination of the three tested genes has worst sensitivity results. The specificity value was 100.0% but the sensitivity value was lower, 93.3% (Table 4).

4.5 Positive predicted value (PPV) and negative predicted value (NPV)

In order to describe the performance of these identification tests the positive predicted value (PPV) and the negative predicted value (NPV) were calculated for each assay and for the combination of assays.

The PPV is important to know the proportion of true positive results of the methodology in test, while the NPV gives the proportion of true negative results (Figure 8).

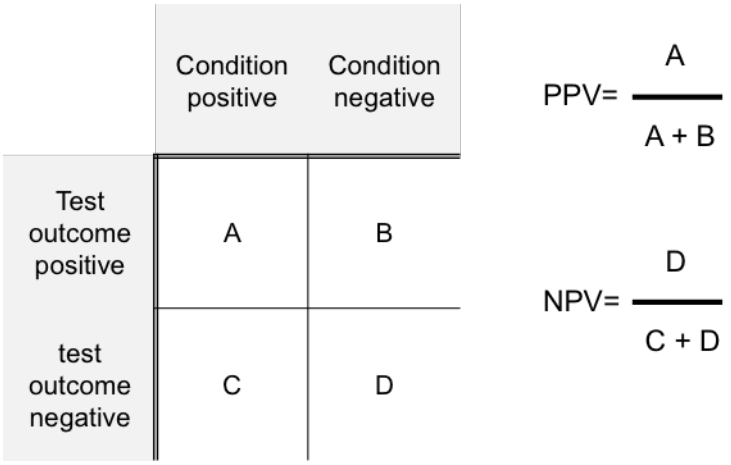

Figure 8. Informative figure that explains the calculation of the positive predicted value and the negative predicted value (adapted from http://sphweb.bumc.bu.edu).

Using the data from the qPCR results (Table 4) for the *lytA*, *piaA* and *hylA* qPCR assays, the PPV is 98.7%, 99.3% and 93.1%, respectively (Table 6 and Supplementary Figure 1). The qPCR assay with the highest probability to have a positive amplification and be *S. pneumoniae* is the *piaA* qPCR. However for the NPV the assay with the highest value is the *lytA* and *hylA* qPCR, with a value of 100.0%. The value for *piaA* qPCR is 92.7% (Table 6). The qPCR assays with the highest probability to be negative when the isolate is not *S. pneumoniae* were the *lytA* and *hylA* qPCR.

When we combined the assays *lytA piaA* qPCR; *lytA hylA* qPCR; *piaA hylA* qPCR and *lytA hylA piaA* qPCR the values for the PPV were: 97.9%, 93.1%, 92.1% and 92.1%, respectively. The values for the NPV were: 92.6%, 100.0%, 92.1% and 92.1%, respectively.

Regarding the PPV and NPV as a single target the *lytA* qPCR have the best values. When we analyse the combined assays the *lytA piaA* qPCR have the best rates of PPV and NPV.

Table 6. PPV and NPV for the single and combined assays.

Target genes	PPV	NPV
lytA	98.7%	100.0%
piaA	99.3%	93.4%
hylA	93.1%	100.0%
lytA piaA	97.9%	92.6%
lytA hylA	93.1%	100.0%
piaA hylA	92.1%	92.1%
lytA piaA hylA	92.1%	92.1%

5. DISCUSSION

Several studies based on a molecular approach that is always dependent of a bacterial culture step have been proposed in the past years. Lull and collaborators proposed a *lytA* RFLP to distinguish *S. pneumoniae* from the closely related species (17). The differentiation was based on specific signatures on the *lytA* present in *S. pneumoniae* that are different from the *lytA* found in the "pneumococcus – like" strains. Nonetheless, more recently, the presence of a homologous of the *lytA* with typical signature has been described in a *S. pseudopneumoniae* strain (38). This finding is important for the sensitivity of this technic and the possible increase of the rate of the false positives. Other targets such as *ply* and *psaA* were proposed for the identification of *S. pneumoniae* (36, 66). However, with only a specificity of 50% for the *ply* gene and the presence of homologous genes in the close relatives turn these two genes in not very good targets for the identification of *S. pneumoniae* (38, 67). The 16S RNA has been proposed as a good target in the identification of *S. pneumoniae*. However the similarity of the *16S RNA* gene sequences between the close relatives is 99%, which is a problem to distinguish between species (80). More recently, Sholz and collaborators (81) described the presence of a highly conserved single nucleotide polymorphism in the *16S RNA* gene sequence that enables the distinction of pneumococcal strains from other species of the Viridans group. Using sequence comparison, or a RFLP technic the authors managed to distinguish almost all *S. pneumoniae* from its relatives. However, this technic as a clinical tool, is time consuming and laborious. The Viridans group MLSA scheme proposed by Bishop and collaborators is a good alternative in clustering at species level a group of isolates (69). However, as a clinical tool it is also laborious, time consuming and expensive (68).

These strategies have a common down side when compared to qPCR. All rely on a first step that is the growth of the bacterium based on culture methods. Therefore there is a decrease in the sensitivity when compared to qPCR. The qPCR technique is able to amplify DNA of non-viable bacteria, increasing the sensitivity (57).

The main aim of this study was to test the best target or targets using qPCR technic for correct identification of *S. pneumoniae*. The genes tested were: *lytA*, as described by Carvalho and collaborators (18); *piaA*, as described by Trzcinski and collaborators (19) and *hylA*, tested for the first time in this study.

To accomplish these aims we used a previously well-characterized collection of isolates (Table 1 and Supplementary Table 4) (51, 55, 82).

In this study we were able to gain insights in to the specificity, sensitivity, PPV and NVP of the three tested genes as targets in a qPCR assay.

In 2007 Carvalho and collaborators described a *lytA* qPCR assay for identification of *S. pneumoniae* (18). Nonetheless, the presence of a homologous *lytA* gene has been described in species of the Viridans group, (*S. pseudopneumoniae* and *S. mitis*) (36-38). In our study we increased the number of isolates of *S. pneumoniae* closely related species, in particular *S. pseudopneumoniae* (44 more isolates, 61 in total), and the specificity and PPV were lower than that obtained by Carvalho and collaborators (18), however the sensitivity and NPV were the same, 100.0% (Table 4 and Table 6). The two *S. pseudopneumoniae* misidentified as *S. pneumoniae* was an unexpectable result. These two strains, 8615 and 9111 were confirmed at species level using the MLSA Viridans streptococci (Figure 6). However, the presence of *lytA* gene has been described in this species (38). Of interest, the *S. pseudopneumoniae* strain described by Shahinas and collaborators (38) to harbor an atypical and a typical *lytA* gene was tested in our study and we obtained no amplification using the qPCR described by Carvalho and collaborators for this strain. Nonetheless if the probe used on the qPCR assay is designed to be specific for *S. pneumoniae* no amplification should occur for other species. This result suggests that the probe and the primers are not totally specific for *S. pneumoniae*, probably interacting with *lytA* homologues different from those described by Shahinas and collaborators, or more remotely, with other part of the genome of these strains. Other hypothesis relies on the fact that the majority of the bacteria of Viridans group is competent, which means that they can acquire exogenous DNA from

species that share the same ecological niche (83, 84). HGT between pneumococcal strains and *S. pseudopneumoniae* has been observed (83-86). Since, the non-pneumococcal strains with positive amplification for *lytA* were collected from sick patients, they are virulent and capable to cause disease, and therefore the acquisition of a functional *lytA* gene could be seen as an advantage.

In this study we focused our attention in a target recently used in qPCR, the *piaA* gene used by Trzcinszki and collaborators (19). The *piaA* gene has been used as a second target when performing the *lytA* qPCR assay to decrease the number of false positives (19).

The *piaA* gene was described to be present only in *S. pneumoniae* and absent in the closely related species (41). However, its presence is not ubiquitous in *S. pneumoniae*, since it is commonly absent in non-typeable pneumococcal strains (41). The specificity and the sensitivity of this gene were not tested until this study. Here we tested for the first time the specificity, sensitivity, PPV and NPV of the *piaA* gene used as a single target.

The value obtained for the specificity and the PPV of the *piaA* qPCR assay was higher than that obtained for the *lytA* qPCR assay (Table 4 and Table 6). However this assay had a lower sensitivity value and NPV than the *lytA* qPCR assay (Table 4 and Table 6). With *piaA* qPCR assay *S. pseudopneumoniae* can be positively identified as *S. pneumoniae* (Table 5 and Figure 6). The *piaA* qPCR assay showed a lower value of sensitivity and NPV when compared with the *lytA* qPCR assay (Table 4 and Table 6). We expected a lower sensitivity since the absence of *piaA* gene in non-typeable pneumococcal strains has been described (41). However, we detected *S. pneumoniae* capsulated strains that were negative for *piaA* qPCR assay (Supplementary Table 4). Such results are not concordant to what has been described so far (41). The presence of a *piaA* homologue gene acquired by HGT could be a possible explanation for the non-pneumococcal positive strain (Table 2). Also, the acquisition of *piaA* gene could be seen as an advantage for the bacteria to survive in iron depletion. However, the lack of *piaA* gene in *S. pneumoniae* capsulated strains could not be seen as advantageous in

colonization and virulence of the *S. pneumoniae* capsulated strains since studies have shown the importance of a functional *piaA* in the virulence of *S. pneumoniae* (28, 87, 88). In addition *piaA* gene is highly conserved at sequence and protein level among *S. pneumoniae* strains (41). Until this study, the presence of the *piaA* gene has not been described in any species of the Viridans group besides *S. pneumoniae* (41). In addition for the first time we did not detect the *piaA* gene in *S. pneumoniae* capsular strains. However, confirmatory tests need to be performed.

After the evaluation of the specificity and the sensitivity of the *piaA* gene, we decided to target a new gene. Since the *hylA* gene encodes for a protein that is considered by some authors a virulence factor (27, 89), and it has been described to be present in *S. pneumoniae* and absent in the closely related species in the Viridans group (46, 77) a qPCR assay was designed to target the *hylA* gene. When compared with the *lytA* and piaA qPCR assays the specificity and PPV for *hylA* qPCR was lower (Table 4 and Table 6). A confirmation at species level was performed using the MLSA Viridans group scheme for all non-pneumococcal positive isolates, and misidentifications between *S. oralis*, *S. pseudopneumoniae* and strains from Viridans group occur when using the *hylA* qPCR (Figure 6). When the alignment of *hylA* gene from the type strain TIGR4 was performed with the NCBI database, no match was found for *S. oralis*, *S. mitis* and *S. pseudopneumoniae* species (Figure 7). Therefore we did not expect any amplification for strains belonging to these species. Nonetheless, the comparison of a sequence with only a database with a limited number of strains available is a limitation and could be seen as a disadvantage for this study design. Furthermore, previous studies showing the specificity of *hylA* gene used genome comparison and only used a low number of isolates (37, 38). Our results indicate that the primers and probe here designed are not specific for *S. pneumoniae* interacting with a possible homologous of the *hylA* gene. This hypothesis can be supported by the acquisition of a *hylA* homologous gene by HGT since the presence of a functional *hylA* gene brings advantages in the colonization and in the ability to cause infection, promoting the virulence of the bacterium by the release or activation of other virulence factors (27, 89).

The *hylA* qPCR assay was capable to correctly identify all the *S. pneumoniae* in the collection with equal values of sensitivity and NPV than those obtained with the *lytA* qPCR assay.

The *lytA* gene as a single target has been used worldwide in qPCR assays. However, in 2013, Trzcinski and collaborators used a second target, *piaA*, in order to decrease the number of false positives (19). In this study we evaluated four parameters: sensitivity, specificity, PPV and NPV. The PPV will give the rate of false positives and therefore in diagnostic tests is the most relevant (90). According to this we have shown that as a single target the *lytA* qPCR assay described by Carvalho and collaborators remains the best choice (Table 4 and Table 6). Nonetheless the best combination of targets in a qPCR assay remains the combination used by Trzcinski and collaborators. Despite the fact that the *lytA hylA* qPCR have the best values of specificity and sensitivity, the PPV and NPV values were worse when compared to *lytA piaA* qPCR (Table 4 and Table 6). No doubt that with the use of two targets the specificity of the assay increases, however in this study we have shown that the PPV decreases with the use of two targets proposed by Trzcinski and collaborators.

In this study we evaluated the use of different C_T (Cycle threshold) values used in the qPCR studies. We propose a standard value in order to compare the results of the studies that use the qPCR assay for the identification of the *S. pneumoniae*. Carvalho and collaborators and Trzcinski and collaborators used for positive identification $C_T<35$ (18, 19). Later, Wyllie and collaborators used as cut off a $C_T<40$ (75). When the results of this study are compared using these two different C_T the best approach seems to use $C_T<35$ as a cut off (Supplementary Table 5). This is supported by the amplification values, obtained for the *S. pneumoniae* strains that were always lower than 33 (Supplementary Table 4). Therefore we propose the use of a $C_T<35$ as standard for all qPCR in the identification of *S. pneumoniae*.

6. CONCLUSIONS

The main goals of this study were to evaluate the specificity, sensitivity, PPV and NPV of the *lytA*, *piaA* and *hylA* genes as targets in a qPCR, and infer what is the better target or, combination of targets, to correctly identify *S. pneumoniae*.

With this study we showed that as a single target, the *lytA* gene seems to be the best target. Nonetheless, *S. pseudopneumoniae* strains could be identified as *S. pneumoniae* based on qPCR for *lytA*. Still, all *S. pneumoniae* strains were correctly identified using *lytA* as a single target.

When two targets are used, the best combination seems to be the *lytA piaA* qPCRs. However, *S. pseudopneumoniae* strains could be identified as *S. pneumoniae*. In addition, some *S. pneumoniae* isolates were not identified by this combination of targets, since negative results for *S. pneumoniae* isolates using *piaA* gene were obtained.

This study showed that the use of the *hylA* gene as a single target or in combination with other targets is not a good solution for the identification of *S. pneumoniae*.

Here we proposed the use of a standard C_T cut-off value of 35 to be used in future studies relying on qPCR for the identification of *S. pneumoniae*.

This study was important to gain insights in to the sensitivity, specificity, PPV and NPV of the qPCR assays currently used by the scientific community in the identification of *S. pneumoniae*.

7. FUTURE PERSPECTIVES

In this study, a large collection of strains was used to evaluate which qPCR *lytA*, *piaA* or *hylA* assay would best identify *S. pneumoniae*. Misidentifications, although rare, were found to occur. Therefore, the focus of future work could be the search for a novel species-specific target that would provide better qPCR results.

There are other topics that I would like to have an answer for. I would like to be able to explain the detection of false negatives for the *hylA* gene, false positives for the *piaA* gene and false positives for the *lytA* gene. In these cases, besides sequencing the complete genes and comparing with reference sequences, one of the options that I would like to study would be the protein activity of HylA, PiaA and LytA, in these isolates, to access if other proteins, with possible differences in the sequence, have the normal activity or a lower/higher activity, influencing the ability of the bacterium to colonize and cause disease. In fact, the answer to these topics could be an approach in a possible future research study.

8. REFERENCES

1. **Breed S. R. MED, Smith N. R.** Bergey's manual of sistematic bacteriology. 7 ed. Baltimore, USA: The Williams & Wilkins Company 1957. p. 507-8.

2. **Facklam R**. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clinical Microbiology Reviews. 2002;15(4):613-30.

3. **Bogaert D**, **De Groot R**, **Hermans PW**. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. The Lancet Infectious Diseases. 2004;4(3):144-54.

4. **Simell B**, **Auranen K**, **Käyhty H**, **Goldblatt D**, **Dagan R**, **O'Brien KL**. The fundamental link between pneumococcal carriage and disease. Expert Review of Vaccines. 2012;11(7):841-55.

5. **Sá-Leão R**, **Nunes S**, **Brito-Avô A**, **Frazão N**, **Simões AS**, **Crisostomo MI**, et al. Changes in pneumococcal serotypes and antibiotypes carried by vaccinated and unvaccinated day-care centre attendees in Portugal, a country with widespread use of the seven-valent pneumococcal conjugate vaccine. Clinical Microbiology and Infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2009;15(11):1002-7.

6. **Almeida ST**, **Nunes S**, **Santos Paulo AC**, **Valadares I**, **Martins S**, **Breia F**, et al. Low prevalence of pneumococcal carriage and high serotype and genotype diversity among adults over 60 years of age living in Portugal. PloS One. 2014;9(3):e90974.

7. **Leiberman A**, **Dagan R**, **Leibovitz E**, **Yagupsky P**, **Fliss DM**. The bacteriology of the nasopharynx in childhood. International Journal of Pediatric Otorhinolaryngology. 1999;49 Suppl 1:S151-3.

8. **Seeley R ST, Tate P**. Anatomia & Fisiologia. 1st ed: Lusodidatica; 1997 1997.

9. **Janssens JP**, **Krause KH**. Pneumonia in the very old. The Lancet Infectious Diseases. 2004;4(2):112-24.

10. **Prevention CDC**. Vaccines and Immunizations [cited 2015 June, 15]. Available from: http://www.cdc.gov/vaccines/vpd-vac/pneumo/vacc-in-short.htm.

11. **National Center for Immunization and Respiratory Diseases DoBD**. Risk factors & transmission: CDC; 2013 [updated June 6, 2013; cited 2015 June 11]. Available from: http://www.cdc.gov/pneumococcal/about/risk-transmission.html.

12. **Almeida ST**, **Nunes S**, **Paulo ACS**, **Valadares I**, **Martins S**, **Breia F**, et al. Low prevalence of pneumococcal carriage and high serotype and genotype diversity among adults over 60 years of age living in Portugal. PloS One. 2014;9(3):e90974.

13. **Liu L**, **Oza S**, **Hogan D**, **Perin J**, **Rudan I**, **Lawn JE**, et al. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet. 2014.

14. **Satzke C**, **Turner P**, **Virolainen-Julkunen A**, **Adrian PV**, **Antonio M**, **Hare KM**, et al. Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*: updated recommendations from the world health organization pneumococcal carriage working group. Vaccine. 2013;32(1):165-79.

15. **Gloria Carvalho M**, **Pimenta FC**, **Jackson D**, **Roundtree A**, **Ahmad Y**, **Millar EV**, et al. Revisiting pneumococcal carriage by use of broth enrichment and PCR techniques for enhanced detection of carriage and serotypes. Journal of Clinical Microbiology. 2010;48(5):1611-8.

16. **Azzari C**, **Moriondo M**, **Indolfi G**, **Cortimiglia M**, **Canessa C**, **Becciolini L**, et al. Realtime PCR is more sensitive than multiplex PCR for diagnosis and serotyping in children with culture negative pneumococcal invasive disease. PloS One. 2010;5(2):e9282.

17. **Llull D**, **Lopez R**, **Garcia E**. Characteristic signatures of the *lytA* gene provide a basis for rapid and reliable diagnosis of *Streptococcus pneumoniae* infections. Journal of Clinical Microbiology. 2006;**44**(4):1250-6.

18. **Carvalho Mda G**, **Tondella ML**, **McCaustland K**, **Weidlich L**, **McGee L**, **Mayer LW**, et al. Evaluation and improvement of real-time PCR assays targeting *lytA*, *ply*, and *psaA* genes for detection of pneumococcal DNA. Journal of Clinical Microbiology. 2007;45(8):2460-6.

19. **Trzciński K**, **Bogaert D**, **Wyllie A**, **Chu ML**, **van der Ende A**, **Bruin JP**, et al. Superiority of trans-oral over trans-nasal sampling in detecting *Streptococcus pneumoniae* colonization in adults. PloS One. 2013;8(3):e60520.

20. **Geno KA**, **Gilbert GL**, **Song JY**, **Skovsted IC**, **Klugman KP**, **Jones C**, et al. Pneumococcal Capsules and Their Types: Past, Present, and Future. Clinical Microbiology Reviews. 2015;28(3):871-99.

21. **Avery OT**, **Dubos R**. The protective action of a specific enzyme against type III pneumococcus infection in mice. The Journal of Experimental Medicine. 1931;54(1):73-89.

22. **Brown EJ**. Interaction of gram-positive microorganisms with complement. Current Topics in Microbiology and Immunology. 1985;121:159-87.

23. **Hardy GG**, **Magee AD**, **Ventura CL**, **Caimano MJ**, **Yother J**. Essential role for cellular phosphoglucomutase in virulence of type 3 *Streptococcus pneumoniae*. Infection and Immunity. 2001;69(4):2309-17.

24. **Magee AD**, **Yother J**. Requirement for capsule in colonization by *Streptococcus pneumoniae*. Infection and Immunity. 2001;69(6):3755-61.

25. **Mitchell TJ**. The pathogenesis of streptococcal infections: from tooth decay to meningitis. Nature Reviews Microbiology. 2003;1(3):219-30.

26. **Harvey RM**, **Hughes CE**, **Paton AW**, **Trappetti C**, **Tweten RK**, **Paton JC**. The impact of pneumolysin on the macrophage response to *Streptococcus pneumoniae* is strain-dependent. PloS One. 2014;9(8):e103625.

27. **Berry AM**, **Paton JC**. Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. Infection and Immunity. 2000;68(1):133-40.

28. **Brown JS**, **Gilliland SM**, **Holden DW**. A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. Molecular Microbiology. 2001;40(3):572-85.

29. **Berry AM**, **Lock RA**, **Thomas SM**, **Rajan DP**, **Hansman D**, **Paton JC**. Cloning and nucleotide sequence of the *Streptococcus pneumoniae* hyaluronidase gene and purification of the enzyme from recombinant *Escherichia coli*. Infection and Immunity. 1994;62(3):1101-8.

30. **Ramos-Sevillano E**, **Urzainqui A**, **Campuzano S**, **Moscoso M**, **González-Camacho F**, **Domenech M**, et al. Pleiotropic effects of cell wall amidase LytA on *Streptococcus pneumoniae* sensitivity to the host immune response. Infection and Immunity. 2015;83(2):591-603.

31. **Rajam G**, **Anderton JM**, **Carlone GM**, **Sampson JS**, **Ades EW**. Pneumococcal surface adhesin A (PsaA): a review. Critical Reviews in Microbiology. 2008;34(3-4):131-42.

32. **Mosser JL**, **Tomasz A**. Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. The Journal of Biological Chemistry. 1970;245(2):287-98.

33. **Martner A**, **Dahlgren C**, **Paton JC**, **Wold AE**. Pneumolysin released during *Streptococcus pneumoniae* autolysis is a potent activator of intracellular oxygen radical production in neutrophils. Infection and Immunity. 2008;76(9):4079-87.

34. **Canvin JR**, **Marvin AP**, **Sivakumaran M**, **Paton JC**, **Boulnois GJ**, **Andrew PW**, et al. The role of pneumolysin and autolysin in the pathology of pneumonia and septicemia in mice infected with a type 2 pneumococcus. The Journal of Infectious Diseases. 1995;172(1):119-23.

35. **Eldholm V**, **Johnsborg O**, **Haugen K**, **Ohnstad HS**, **Hàvarstein LS**. Fratricide in *Streptococcus pneumoniae*: contributions and role of the cell wall hydrolases CbpD, LytA and LytC. Microbiology. 2009;155(Pt 7):2223-34.

36. **Whatmore AM**, **Efstratiou A**, **Pickerill AP**, **Broughton K**, **Woodard G**, **Sturgeon D**, et al. Genetic relationships between clinical isolates of *Streptococcus pneumoniae*, *Streptococcus oralis*, and *Streptococcus mitis*: characterization of "Atypical" pneumococci and organisms allied to *S. mitis* harboring *S. pneumoniae* virulence factor-encoding genes. Infection and Immunity. 2000;68(3):1374-82.

37. **Denapaite D**, **Brückner R**, **Nuhn M**, **Reichmann P**, **Henrich B**, **Maurer P**, et al. The genome of *Streptococcus mitis* B6--what is a commensal? PloS One. 2010;5(2):e9426.

38. **Shahinas D**, **Thornton CS**, **Tamber GS**, **Arya G**, **Wong A**, **Jamieson FB**, et al. Comparative genomic analyses of *Streptococcus pseudopneumonae* provide insight into virulence and commensalism dynamics. PloS One. 2013;8(6):e65670.

39. **Sutcliffe IC**, **Russell RR**. Lipoproteins of gram-positive bacteria. Journal of Bacteriology. 1995;177(5):1123-8.

40. **Tai SS**, **Yu C**, **Lee JK**. A solute binding protein of *Streptococcus pneumoniae* iron transport. FEMS Microbiology Letters. 2003;220(2):303-8.

41. **Whalan RH**, **Funnell SG**, **Bowler LD**, **Hudson MJ**, **Robinson A**, **Dowson CG**. Distribution and genetic diversity of the ABC transporter lipoproteins PiuA and PiaA within *Streptococcus pneumoniae* and related streptococci. Journal of Bacteriology. 2006;188(3):1031-8.

42. **Boulnois GJ**. Pneumococcal proteins and the pathogenesis of disease caused by *Streptococcus pneumoniae*. Journal of General Microbiology. 1992;138(2):249-59.

43. **Allen AG**, **Lindsay H**, **Seilly D**, **Bolitho S**, **Peters SE**, **Maskell DJ**. Identification and characterisation of hyaluronate lyase from *Streptococcus suis*. Microbial Pathogenesis. 2004;36(6):327-35.

44. **Linker A**, **Mayer K**. Production of unsaturated uronides by bacterial hyaluronidases. Nature. 1954;174(4443):1192-3.

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

46. **Rolland K**, **Marois C**, **Siquier V**, **Cattier B**, **Quentin R**. Genetic features of *Streptococcus agalactiae* strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and *hylB* gene analysis. Journal of Clinical Microbiology. 1999;37(6):1892-8.

47. **Donati C**, **Hiller NL**, **Tettelin H**, **Muzzi A**, **Croucher NJ**, **Angiuoli SV**, et al. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. Genome Biology. 2010;**11**(10):R107.

48. **Fraser C**, **Hanage WP**, **Spratt BG**. Recombination and the nature of bacterial speciation. Science. 2007;**315**(5811):476-80.

49. **Rolo D**, **Simões AS**, **Domenech A**, **Fenoll A**, **Liñares J**, **de Lencastre H**, et al. Disease isolates of *Streptococcus pseudopneumoniae* and non-typeable *S. pneumoniae* presumptively identified as atypical *S. pneumoniae* in Spain. PloS One. 2013;8(2):e57047.

50. **Arbique JC**, **Poyart C**, **Trieu-Cuot P**, **Quesne G**, **Carvalho Mda G**, **Steigerwalt AG**, et al. Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus*

pseudopneumoniae sp. nov. Journal of Clinical Microbiology. 2004;**42**(10):4686-96.

51. **Rolo D**, **A SS**, **Domenech A**, **Fenoll A**, **Liñares J**, **de Lencastre H**, et al. Disease isolates of *Streptococcus pseudopneumoniae* and non-typeable *S. pneumoniae* presumptively identified as atypical *S. pneumoniae* in Spain. PloS One. 2013;**8**(2):e57047.

52. **Harf-Monteil C**, **Granello C**, **Le Brun C**, **Monteil H**, **Riegel P**. Incidence and pathogenic effect of *Streptococcus pseudopneumoniae*. Journal of Clinical Microbiology. 2006;**44**(6):2240-1.

53. **Doern CD**, **Burnham CA**. It's not easy being green: the viridans group streptococci, with a focus on pediatric clinical manifestations. Journal of Clinical Microbiology. 2010;**48**(11):3829-35.

54. **Douglas CW**, **Heath J**, **Hampton KK**, **Preston FE**. Identity of Viridans streptococci isolated from cases of infective endocarditis. Journal of Medical Microbiology. 1993;39(3):179-82.

55. **Simões AS**, **Sá-Leão R**, **Eleveld MJ**, **Tavares DA**, **Carriço JA**, **Bootsma HJ**, et al. Highly penicillin-resistant multidrug-resistant pneumococcus-like strains colonizing children in Oeiras, Portugal: genomic characteristics and implications for surveillance. Journal of Clinical Microbiology. 2010;48(1):238-46.

56. **Kilian M**, **Poulsen K**, **Blomqvist T**, **Hävarstein LS**, **Bek-Thomsen M**, **Tettelin H**, et al. Evolution of *Streptococcus pneumoniae* and its close commensal relatives. PloS One. 2008;**3**(7):e2683.

57. **Albrich WC**, **Madhi SA**, **Adrian PV**, **van Niekerk N**, **Mareletsi T**, **Cutland C**, et al. Use of a rapid test of pneumococcal colonization density to diagnose pneumococcal pneumonia. Clinical Infectious Diseases : an official publication of the Infectious Diseases Society of America. 2012;54(5):601-9.

58. **M.P. CHL**. Microbiological methods. 5 ed: Butterworths&CO; 1984. p. 110.

59. **Mundy LS**, **Janoff EN**, **Schwebke KE**, **Shanholtzer CJ**, **Willard KE**. Ambiguity in the identification of *Streptococcus pneumoniae*. Optochin, bile solubility, quellung, and the AccuProbe DNA probe tests. American Journal of Clinical Pathology. 1998;109(1):55-61.

60. **Ikryannikova LN**, **Lapin KN**, **Malakhova MV**, **Filimonova AV**, **Ilina EN**, **Dubovickaya VA**, et al. Misidentification of alpha-hemolytic streptococci by routine tests in clinical practice. Infection, genetics and evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases. 2011;11(7):1709-15.

61. **Nunes S**, **Sá-Leão R**, **de Lencastre H**. Optochin resistance among *Streptococcus pneumoniae* strains colonizing healthy children in Portugal. Journal of Clinical Microbiology. 2008;46(1):321-4.

62. **Rouff K. L. R. A. W. BD**. Manual of Clinical Microbiology. 7 ed. Washinton DC: American Society for Microbiology; 1999. p. 283-96.

63. **Obregón V**, **García P**, **García E**, **Fenoll A**, **Lopez R**, **Garcia JL**. Molecular peculiarities of the *lytA* gene isolated from clinical pneumococcal strains that are bile insoluble. Journal of Clinical Microbiology. 2002;40(7):2545-54.

64. **Lund E**. Laboratory diagnosis of Pneumococcus infections. Bulletin of the World Health Organization. 1960;23:5-13.

65. **Holmberg H**, **Danielsson D**, **Hardie J**, **Krook A**, **Whiley R**. Cross-reactions between alpha-streptococci and omniserum, a polyvalent pneumococcal serum, demonstrated by direct immunofluorescence, immunoelectroosmophoresis, and latex agglutination. Journal of Clinical Microbiology. 1985;21(5):745-8.

66. **Messmer TO**, **Sampson JS**, **Stinson A**, **Wong B**, **Carlone GM**, **Facklam RR**. Comparison of four polymerase chain reaction assays for specificity in the identification of *Streptococcus pneumoniae*. Diagnostic Microbiology and Infectious Disease. 2004;49(4):249-54.

67. **Jado I**, **Fenoll A**, **Casal J**, **Pérez A**. Identification of the *psaA* gene, coding for pneumococcal surface adhesin A, in viridans group streptococci other than *Streptococcus pneumoniae*. Clinical and Diagnostic Laboratory Immunology. 2001;8(5):895-8.

68. **Hanage WP**, **Fraser C**, **Spratt BG**. Sequences, sequence clusters and bacterial species. Philosophical transactions of the Royal Society of London Series B, Biological Sciences. 2006;361(1475):1917-27.

69. **Bishop CJ**, **Aanensen DM**, **Jordan GE**, **Kilian M**, **Hanage WP**, **Spratt BG**. Assigning strains to bacterial species via the internet. BMC Biology. 2009;7:3.

70. **Hoshino T**, **Fujiwara T**, **Kilian M**. Use of phylogenetic and phenotypic analyses to identify nonhemolytic streptococci isolated from bacteremic patients. Journal of Clinical Microbiology. 2005;43(12):6073-85.

71. **Botstein D**, **White RL**, **Skolnick M**, **Davis RW**. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics. 1980;32(3):314-31.

72. **Nathans D**, **Smith HO**. Restriction endonucleases in the analysis and restructuring of dna molecules. Annual Review of Biochemistry. 1975;44:273-93.

73. **Bustin SA**, **Benes V**, **Garson JA**, **Hellemans J**, **Huggett J**, **Kubista M**, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry. 2009;55(4):611-22.

74. **Kim DW**, **Kilgore PE**, **Kim EJ**, **Kim SA**, **Anh DD**, **Dong BQ**, et al. The enhanced pneumococcal LAMP assay: a clinical tool for the diagnosis of meningitis due to *Streptococcus pneumoniae*. PloS One. 2012;7(8):e42954.

75. **Wyllie AL**, **Chu ML**, **Schellens MH**, **van Engelsdorp Gastelaars J**, **Jansen MD**, **van der Ende A**, et al. *Streptococcus pneumoniae* in saliva of dutch primary school children. PloS One. 2014;9(7):e102045.

76. **Hajia M**, **Farzanehkhah M**, **Hajihashemi B**, **Dolatyar A**, **Imani M**, **Saburian R**, et al. Real-Time assay as a tool for detecting *lytA* gene in S*treptococcus pneumoniae* isolates. Cell Journal. 2014;16(2):141-6.

77. **Madhour A**, **Maurer P**, **Hakenbeck R**. Cell surface proteins in *S. pneumoniae, S. mitis and S. oralis*. Iranian Journal of Microbiology. 2011;3(2):58-67.

78. **M.P CCHL**. Microbiological Methods. 5 ed: Butterworth&CO; 1984.

79. **Burns M**, **Valdivia H**. Modelling the limit of detection in real-time quantitative PCR. European Food Research and Technology. 2008;226(6):1513-24.

80. **Kawamura Y**, **Hou XG**, **Sultana F**, **Miura H**, **Ezaki T**. Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. International Journal of Systematic Bacteriology. 1995;45(2):406-8.

81. **Scholz CF**, **Poulsen K**, **Kilian M**. Novel molecular method for identification of *Streptococcus pneumoniae* applicable to clinical microbiology and 16S rRNA sequence-based microbiome studies. Journal of Clinical Microbiology. 2012;50(6):1968-73.

82. **Simões AS**, **Valente C**, **de Lencastre H**, **Sá-Leão R**. Rapid identification of noncapsulated *Streptococcus pneumoniae* in nasopharyngeal samples allowing detection of co-colonization and reevaluation of prevalence. Diagnostic Microbiology and Infectious Disease. 2011;71(3):208-16.

83. **Claverys JP**, **Martin B**, **Polard P**. The genetic transformation machinery: composition, localization, and mechanism. FEMS Microbiology Reviews. 2009;33(3):643-56.

84. **Johnsborg O**, **Håvarstein LS**. Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae*. FEMS Microbiology Reviews. 2009;33(3):627-42.

85. **Gogarten JP**, **Townsend JP**. Horizontal gene transfer, genome innovation and evolution. Nature Reviews Microbiology. 2005;3(9):679-87.

86. **Majewski J**, **Zawadzki P**, **Pickerill P**, **Cohan FM**, **Dowson CG**. Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. Journal of bacteriology. 2000;182(4):1016-23.

87. **Brown JS, Ogunniyi AD, Woodrow MC, Holden DW, Paton JC**. Immunization with components of two iron uptake ABC transporters protects mice against systemic *Streptococcus pneumoniae* infection. Infection and Immunity. 2001;69(11):6702-6.

88. **Jomaa M**, **Yuste J**, **Paton JC**, **Jones C**, **Dougan G**, **Brown JS**. Antibodies to the iron uptake ABC transporter lipoproteins PiaA and PiuA promote opsonophagocytosis of *Streptococcus pneumoniae*. Infection and Immunity. 2005;73(10):6852-9.

89. **Homer K**, **Shain H**, **Beighton D**. The role of hyaluronidase in growth of *Streptococcus intermedius* on hyaluronate. Advances in Experimental Medicine and Biology. 1997;418:681-3.

90. **Altman DG**, **Bland JM**. Diagnostic tests 2: Predictive values. BMJ (Clinical Research ed). 1994;309(6947):102.

9. APPENDIX

9.1 Supplemental data

Supplementary Table 1. Sequences of primers and probes used for the qPCR assays.

Gene	Primers and probes sequence	Ref.
lytA	*lytA*_F: 5' – ACGCAATCTAGCAGATGAAGCA - 3' *lytA*_R: 5' – TCGTGCGTTTTAATTCCAGCT - 3' *lytA*_P:5' 6FAM-TGCCGAAAACGCXTTGATACAGGGAG – PH - 3'	18
piaA	*piaA*_F: 5'-CATTGGTGGCTTAGTAAGTGCAA - 3' *piaA*_R: 5'- TACTAACACAAGTTCCTGATAAGGCAAGT - 3' *piaA*_P: 5'- ROX-TGTAAGCGGAAAAGCAGGCCTTACCC – BBQ - 3'	19
hylA	*hylA*_F: 5'-CCAAATTAAACCAGGAATTGGA-3' *hylA*_R: 5'- CTCCAATTTCCGATAAGTGGCA-3. *hylA*_P: 5' Cy5 – TCAAGTCAGGCGGACCGCA – BBQ - 3'	This study

Supplementary Table 2. Primers and respective sequence used on the amplification of the eight housekeeping genes in the MLSA method.

Primer	Sequence	Annealing temperatures (°C)	Ref
map-up	5' - GCWGACTCWTGTTGGGCWTATGC - 3'	55	(69)
map-dn	5' - TTARTAAGTTCYTTCTTCDCCTTG - 3'		(69)
pfl-up	5' - AACGTTGCTTACTCTAAACAAACTGG - 3'	55	(69)
pfl-dn	5' - ACTTCRTGGAAGACACGTTGWGTC - 3'		(69)
ppaC-up	5' - GACCAYAATGAATTYCARCAATC - 3'	50	(69)
ppaC-dn	5' - TGAGGNACMACTTGTTTSTTACG - 3'		(69)
pyk-up	5' - GCGGTWGAAWTCCGTGGTG - 3'	50	(69)
pyk-dn	5' - GCAAGWGCTGGGAAAGGAAT - 3'		(69)
rpoB-up	5' - AARYTIGGMCCTGAAGAAAT - 3'	50	(69)
rpoB-dn	5' - TGIARTTTRTCATCAACCATGTG - 3'		(69)
sodA-up	5' - TRCAYCATGAYAARCACCAT - 3'	50	(69)
sodA-dn	5' - ARRTARTAMGCRTGYTCCCARACRTC - 3'		study
tuf-up	5' - GTTGAAATGGAAATCCGTGACC - 3'	55	(69)
tuf-dn	5' - GTTGAAGAATGGAGTGTGACG - 3'		(69)

Supplementary Table 3. Primers and probe initially designed for the *hylA* qPCR.

Gene	Primers and probes sequence
	_hylA_1F:5' – CCAAATTAAACCAGGAATTGGA - 3'
	_hylA_1R:5' – CTCCAATTTCCGATAAGTGGCA – 3'.
hylA	*hylA*_2F:5' – ATTGGAAGGAAAGGTAGCTGAT - 3'
	*hylA*_2R:5' – AGACGTTTAGACTGACGGTGA – 3'.
	_hylA_P :5' Cy5 – TCAAGTCAGGCGGACCGCA - BBQ 3'

The primers and the probe used in the *hylA* qPCR are indicated in bold.

Supplementary Table 4. Study collection – characterization.

Strain	Species	Ser.	Average C_T for *lytA*	Average C_T for *piaA*	Average C_T for *hylA*	R
37	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
281	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
337	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
531	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
603	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
939	*S. pseudopneumoniae*	-	39	26	NA	(51)
1137	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
1173	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
1304	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
1324	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
1544	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
1752	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
1848	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
1850	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
1927	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
2161	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
2353	*S. pseudopneumoniae*	-	39	NA	NA	(51)
2504	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
2522	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
2565	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
2581	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
2597	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
2609	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
2615	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
2621	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
3075	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
3194	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
3205	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
3473	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
3738	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
4526	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
6265	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
6338	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
6339	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
6408	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
6486	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
6669	*S. pseudopneumoniae*	-	NA	NA	NA	(51)

Ser.- Serotype; R – Reference; NA- No amplification; Bold – Invasive strain.

Supplementary Table 4. Study collection – characterization (Cont.).

Strain	Species	Ser.	Average C_T for *lytA*	Average C_T for *piaA*	Average C_T for *hylA*	R
6744	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
6787	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
7253	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
7327	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
7332	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
7842	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
7943	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
8615	*S. pseudopneumoniae*	-	23	NA	24	(51)
8646	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
8937	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
8971	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9012	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9013	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9015	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9096	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9111	*S. pseudopneumoniae*	-	33	NA	35	(51)
9230	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9275	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9545	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9731	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9781	*S. pseudopneumoniae*	-	NA	NA	NA	(51)
9786	*S. pseudopneumoniae*	-	37	NA	NA	(51)
1964	Viridans group	-	NA	NA	NA	(51)
6147	Viridans group	-	NA	NA	NA	(51)
6760	Viridans group	-	NA	NA	25	(51)
7728	Viridans group	-	NA	NA	NA	(51)
7755	Viridans group	-	NA	NA	NA	(51)
8271	Viridans group	-	NA	NA	39	(51)
8277	Viridans group	-	NA	NA	NA	(51)
8447	Viridans group	-	NA	NA	25	(51)
8592	Viridans group	-	NA	NA	21	(51)
8692	Viridans group	-	NA	NA	23	(51)
8943	Viridans group	-	NA	NA	NA	(51)
9279	Viridans group	-	NA	NA	39	(51)
8482	Viridans group	-	NA	NA	37	(51)
7693	Viridans group	-	NA	NA	35	(51)
PT5274	Viridans group	-	NA	NA	NA	(55)

Supplementary Table 4. Study collection – characterization (Cont.).

Strain	Species	Ser.	Average C_T for *lytA*	Average C_T for *piaA*	Average C_T for *hylA*	R
PT5283	Viridans group	-	NA	NA	NA	(55)
PT5295b	Viridans group	-	39	NA	NA	(55)
PT5346b	Viridans group	-	NA	NA	25	(55)
PT5525b	Viridans group	-	NA	NA	NA	(55)
PT5525c	Viridans group	-	NA	NA	NA	(55)
PT5532	Viridans group	-	NA	NA	NA	(55)
PT5534b	Viridans group	-	NA	NA	NA	(55)
PT5557b	Viridans group	-	38	NA	NA	(55)
PT5590a	Viridans group	-	NA	NA	25	(55)
PT5590b	Viridans group	-	NA	NA	NA	(55)
PT5645a	Viridans group	-	NA	NA	NA	(55)
PT5645b	Viridans group	-	NA	NA	NA	(55)
PT5714	Viridans group	-	NA	NA	NA	(55)
PT5729	Viridans group	-	NA	NA	NA	(55)
PT5736b	Viridans group	-	NA	NA	NA	(55)
PT5779	Viridans group	-	40	NA	NA	(55)
PT5787b	Viridans group	-	NA	NA	NA	(55)
PT5790b	Viridans group	-	NA	NA	NA	(55)
PT5793b	Viridans group	-	NA	NA	NA	(55)
PT5794b	Viridans group	-	NA	NA	NA	(55)
PT5796b	Viridans group	-	NA	NA	NA	(55)
PT5798b	Viridans group	-	NA	NA	33	(55)
PT5804	Viridans group	-	NA	NA	NA	(55)
DSM20480	*S. bovis*	-	NA	NA	NA	(82)
DSM20066	*S. oralis*	-	NA	NA	NA	(82)
DSMZ6784	*S. agalactiae*	-	NA	NA	NA	(82)
DSMZ20563	*S. anginosus*	-	NA	NA	NA	(82)
DSM20715	*S. canis*	-	39	NA	NA	(82)
DSMZ20662	*S. dysgalactiae sub. disgalactiae*	-	NA	NA	NA	(82)
DSMZ6176	*S. dysgalactiae sub. equisimilis*	-	NA	NA	NA	(82)
DSMZ20727	*S. equi sub. zooepidemicus*	-	NA	NA	NA	(82)
DSM20568	*S. gordini*	-	NA	NA	NA	(82)
DSM6777	*S. gordini*	-	NA	NA	NA	(82)
DSM12492	*S. infantis*	-	NA	NA	NA	(82)
DSM12643	*S. mitis*	-	NA	NA	NA	(82)
DSM20523	*S. mutans*	-	NA	NA	NA	(82)
DSM20627	*S. oralis*	-	NA	NA	NA	(82)

Supplementary Table 4. Study collection – characterization (Cont.).

Strain	Species	Ser.	Average C_T for *lytA*	Average C_T for *piaA*	Average C_T for *hylA*	R
DSM6778	*S. parasanguinis*	-	NA	NA	NA	(82)
DSM12493	*S. peroris*	-	NA	NA	NA	(82)
DSM20565	*S. pyogenes*	-	NA	NA	NA	(82)
DSM20560	*S. salivarius*	-	NA	NA	NA	(82)
DSM20567	*S. sanguinis*	-	NA	NA	NA	(82)
DSM14990	*S. sinensis*	-	NA	NA	NA	(82)
DSM5636	*S. vestibularis*	-	NA	NA	NA	(82)
DSM20379	*S. oralis* (*S. viridans* I)	-	NA	NA	NA	(82)
DSM20395	*S. oralis* (*S. viridans II*)	-	NA	NA	NA	(82)
DSM20392	*Strep. spp* (*S. viridans* III)	-	NA	NA	NA	(82)
DSM20377	*Strep. spp* (*S. viridans* IV)	-	NA	NA	NA	(82)
DSM8249	*S. cristatus*	-	NA	NA	NA	(82)
NCTC10389	*S. equinis*	-	NA	NA	NA	(82)
NCTC11324	*S. intermedius*	-	NA	NA	NA	(82)
NCTC11325	*S. constelatus*	-	NA	NA	NA	(82)
IS7943	*S. pseudopneumoniae*	-	NA	NA	NA	(38)
PT5479	*S. pseudopneumoniae*	-	NA	NA	NA	(82)
DCC 1365	*S. pneumoniae*	23F	21	37	24	(82)
DCC 635	*S. pneumoniae*	NT	23	NA	23	(82)
DCC 646	*S. pneumoniae*	23F	23	NA	24	(82)
PT1214	*S. pneumoniae*	4	21	21	21	(82)
PT1263	*S. pneumoniae*	3	22	23	23	(82)
PT1265	*S. pneumoniae*	19F	22	24	22	(82)
PT1274	*S. pneumoniae*	31	22	26	25	(82)
PT1283	*S. pneumoniae*	23F	20	22	20	(82)
PT1314	*S. pneumoniae*	35F	23	27	24	(82)
PT1333	*S. pneumoniae*	18A	21	23	22	(82)
PT1348	*S. pneumoniae*	18C	19	23	21	(82)
PT1493	*S. pneumoniae*	NT	20	24	22	(82)
PT1550	*S. pneumoniae*	15F	21	25	22	(82)
PT1581	*S. pneumoniae*	31	21	23	22	(82)
PT1592	*S. pneumoniae*	23F	22	23	22	(82)
PT1683	*S. pneumoniae*	NT	21	22	21	(82)
PT1718	*S. pneumoniae*	NT	22	NA	25	(82)
PT1721	*S. pneumoniae*	24F	21	21	21	(82)
PT1730	*S. pneumoniae*	15A	22	24	22	(82)
PT1804b	*S. pneumoniae*	NT	22	NA	23	(82)

Supplementary Table 4. Study collection – characterization (Cont.).

Strain	Species	Ser.	Average C_T for *lytA*	Average C_T for *piaA*	Average C_T for *hylA*	R
PT1809	*S. pneumoniae*	19F	21	24	23	(82)
PT2111	*S. pneumoniae*	15B	20	24	21	(82)
PT2293b	*S. pneumoniae*	NT	21	NA	22	(82)
PT2398	*S. pneumoniae*	35F	21	23	25	(82)
PT2436	*S. pneumoniae*	19A	21	24	22	(82)
PT2445a	*S. pneumoniae*	11A	19	21	20	(82)
PT2537	*S. pneumoniae*	4	20	22	21	(82)
PT2548	*S. pneumoniae*	15C	21	22	22	(82)
PT2593	*S. pneumoniae*	15F	21	22	22	(82)
PT2605	*S. pneumoniae*	18A	21	26	24	(82)
PT2655	*S. pneumoniae*	33F	20	21	23	(82)
PT2808	*S. pneumoniae*	19A	21	21	21	(82)
PT2942	*S. pneumoniae*	24F	21	21	21	(82)
PT3104	*S. pneumoniae*	6B	20	25	23	(82)
PT3201	*S. pneumoniae*	NT	21	26	24	(82)
PT3341	*S. pneumoniae*	34	20	21	20	(82)
PT3354	*S. pneumoniae*	7F	21	22	20	(82)
PT3400	*S. pneumoniae*	9N	19	25	21	(82)
PT3491	*S. pneumoniae*	35F	20	20	26	(82)
PT3501	*S. pneumoniae*	37	20	24	21	(82)
PT3625a	*S. pneumoniae*	38	21	22	24	(82)
PT3625B	*S. pneumoniae*	7A	23	26	25	(82)
PT3626	*S. pneumoniae*	14	21	23	23	(82)
PT3700	*S. pneumoniae*	10	20	22	21	(82)
PT3919	*S. pneumoniae*	6B	22	24	22	(82)
PT4014	*S. pneumoniae*	NT	20	26	23	(82)
PT4071	*S. pneumoniae*	22F	21	23	21	(82)
PT4076	*S. pneumoniae*	3	21	21	22	(82)
PT4140	*S. pneumoniae*	9V	20	22	24	(82)
PT4188	*S. pneumoniae*	18C	21	21	21	(82)
PT4216	*S. pneumoniae*	7F	21	23	21	(82)
PT4217	*S. pneumoniae*	8	19	23	22	(82)
PT4222	*S. pneumoniae*	NT	20	26	22	(82)
PT4232	*S. pneumoniae*	16F	20	24	22	(82)
PT4264	*S. pneumoniae*	39	20	26	23	(82)
PT4272	*S. pneumoniae*	15B	19	20	19	(82)
PT4400a	*S. pneumoniae*	23F	19	21	19	(82)

Supplementary Table 4. Study collection – characterization (Cont.).

Strain	Species	Ser.	Average C_T for *lytA*	Average C_T for *piaA*	Average C_T for *hylA*	R
PT4481b	*S. pneumoniae*	35F	20	21	21	(82)
PT449	*S. pneumoniae*	6A	21	22	22	(82)
PT450	*S. pneumoniae*	10A	20	22	22	(82)
PT460	*S. pneumoniae*	29	20	24	24	(82)
PT4678b	*S. pneumoniae*	23F	24	25	24	(82)
PT4733	*S. pneumoniae*	29	20	25	23	(82)
PT4737	*S. pneumoniae*	10	22	24	24	(82)
PT476	*S. pneumoniae*	11A	20	21	20	(82)
PT4912	*S. pneumoniae*	38	21	22	21	(82)
PT4928	*S. pneumoniae*	11A	18	21	19	(82)
PT4973	*S. pneumoniae*	9A	23	30	28	(82)
PT500	*S. pneumoniae*	3	23	26	25	(82)
PT5021	*S. pneumoniae*	23A	23	26	25	(82)
PT5037	*S. pneumoniae*	1	20	22	21	(82)
PT5052B	*S. pneumoniae*	18F	20	25	23	(82)
PT5066	*S. pneumoniae*	42	21	22	27	(82)
PT5082B	*S. pneumoniae*	18F	21	25	22	(82)
PT5137	*S. pneumoniae*	20	21	22	23	(82)
PT5161a	*S. pneumoniae*	14	21	21	21	(82)
PT5161b	*S. pneumoniae*	3	24	24	24	(82)
PT526b	*S. pneumoniae*	11A	23	39	23	(82)
PT829	*S. pneumoniae*	23B	21	22	23	(82)
PT952	*S. pneumoniae*	14	21	21	22	(82)
WL 10	*S. pneumoniae*	9V	21	25	23	(82)
WL 1018	*S. pneumoniae*	1	21	21	22	(82)
WL 1055	*S. pneumoniae*	10A	19	21	22	(82)
WL 1059	*S. pneumoniae*	15A	21	21	22	(82)
WL 1084	*S. pneumoniae*	NT	21	37	22	(82)
WL 1185	*S. pneumoniae*	37	21	22	24	(82)
WL 1200	*S. pneumoniae*	24B	21	21	21	(82)
WL 1202	*S. pneumoniae*	19F	21	24	22	(82)
WL 1215	*S. pneumoniae*	42	20	25	25	(82)
WL 1223a	*S. pneumoniae*	19F	23	28	26	(82)
WL 1259	*S. pneumoniae*	33B	19	21	20	(82)
WL 1281	*S. pneumoniae*	15C	20	21	20	(82)
WL 1357	*S. pneumoniae*	9N	20	21	21	(82)
WL 1375	*S. pneumoniae*	19F	20	20	22	(82)

Supplementary Table 4. Study collection – characterization (Cont.).

Strain	Species	Ser.	Average C_T for *lytA*	Average C_T for *piaA*	Average C_T for *hylA*	R
WL 1387	*S. pneumoniae*	21	20	21	22	(82)
WL 1402	*S. pneumoniae*	34	20	24	22	(82)
WL 1471	*S. pneumoniae*	38	20	22	20	(82)
WL 1472	*S. pneumoniae*	14	21	21	22	(82)
WL 1498	*S. pneumoniae*	9A	21	21	21	(82)
WL 1514	*S. pneumoniae*	NT	23	NA	23	(82)
WL 1542	*S. pneumoniae*	16F	21	21	20	(82)
WL 183	*S. pneumoniae*	23A	20	21	20	(82)
WL 186	*S. pneumoniae*	17	20	23	21	(82)
WL 22	*S. pneumoniae*	17	21	22	22	(82)
WL 235	*S. pneumoniae*	38	19	22	21	(82)
WL 253.1	*S. pneumoniae*	16F	21	21	21	(82)
WL 308	*S. pneumoniae*	18B	21	21	20	(82)
WL 386	*S. pneumoniae*	33F	21	22	22	(82)
WL 480	*S. pneumoniae*	38	21	23	22	(82)
WL 399	*S. pneumoniae*	23B	20	23	21	(82)
WL 413	*S. pneumoniae*	38	22	26	24	(82)
WL 418	*S. pneumoniae*	18B	21	23	21	(82)
WL 555	*S. pneumoniae*	11A	20	23	21	(82)
WL 560.1	*S. pneumoniae*	16F	20	23	21	(82)
WL 586	*S. pneumoniae*	12B	19	22	21	(82)
WL 594	*S. pneumoniae*	12A	20	22	20	(82)
WL 70	*S. pneumoniae*	6A	20	22	20	(82)
WL 737	*S. pneumoniae*	12F	20	22	21	(82)
WL 746	*S. pneumoniae*	9L	20	21	22	(82)
WL 762	*S. pneumoniae*	9L	19	22	20	(82)
WL 777	*S. pneumoniae*	12F	19	21	20	(82)
WL 86.1	*S. pneumoniae*	21	20	22	21	(82)
WL 920	*S. pneumoniae*	22F	18	21	19	(82)
Spain23F	*S. pneumoniae*	23F	21	22	23	(82)
Spain6B	*S. pneumoniae*	6B	21	24	22	(82)
Spain9V	*S. pneumoniae*	9V	20	22	21	(82)
PMEN1 Spain23F	*S. pneumoniae*	23F	21	22	23	(82)
PMEN2 Spain6B	*S. pneumoniae*	6B	21	24	22	(82)
PMEN3 Spain9V	*S. pneumoniae*	9V	20	22	21	(82)
PMEN4 Tennessee23F	*S. pneumoniae*	23F	21	21	23	(82)
PMEN5 Spain14	*S. pneumoniae*	14	21	25	23	(82)

Supplementary Table 4. Study collection – characterization (Cont.).

Strain	Species	Ser.	Average C_T for *lytA*	Average C_T for *piaA*	Average C_T for *hylA*	R
PMEN6 Hungary19A	*S. pneumoniae*	19A	21	24	22	(82)
PMEN7 S. Africa19A	*S. pneumoniae*	19A	21	22	21	(82)
PMEN8 S. Africa6B	*S. pneumoniae*	6B	19	22	21	(82)
PMEN9 England14	*S. pneumoniae*	14	20	21	22	(82)
PMEN10 CSR14	*S. pneumoniae*	14	21	23	21	(82)
PMEN11 CSR19A	*S. pneumoniae*	19A	20	21	20	(82)
PMEN12 Finland6B	*S. pneumoniae*	6B	21	25	22	(82)
PMEN13 S. Africa19A	*S. pneumoniae*	19A	19	21	19	(82)
PMEN14 Taiwan19F	*S. pneumoniae*	19F	20	22	20	(82)
PMEN15 Taiwan23F	*S. pneumoniae*	23F	20	23	21	(82)
PMEN16 Poland23F	*S. pneumoniae*	23F	22	26	23	(82)
PMEN17 Maryland6B	*S. pneumoniae*	6B	20	39	23	(82)
PMEN18 Tennessee14	*S. pneumoniae*	14	20	23	21	(82)
PMEN19 Colombia5	*S. pneumoniae*	5	21	23	21	(82)
PMEN20 Poland6B	*S. pneumoniae*	6B	22	23	25	(82)
PMEN21 Portugal19F	*S. pneumoniae*	19F	20	22	20	(82)
PMEN22 Greece6B	*S. pneumoniae*	6B	21	23	21	(82)
PMEN23 N. Carolina6A	*S. pneumoniae*	6A	20	23	21	(82)
PMEN24 Utah35B	*S. pneumoniae*	35B	23	25	23	(82)
PMEN25 Sweden15A	*S. pneumoniae*	15A	20	22	21	(82)
PMEN26 Colombia23F	*S. pneumoniae*	23F	22	25	24	(82)
PMEN41 Portugal6A	*S. pneumoniae*	6A	20	22	20	(82)

Ser.- Serotype; R – Reference; NA- No amplification; Bold – Invasive strain.

Supplementary Table 5. PPV and NPV for the single and combined assays tested with a cut off C_T<35 and C_T<40.

Target genes	PPV C_T<35	PPV C_T<40	NPV C_T<35	NPV C_T<40
lytA	98.7%	95.5%	100.0%	100.0%
piA	99.3%	99.3%	92.7%	95.5%
hylA	93.1%	90.9%	100.0%	100.0%
lytA piaA	97.9%	94.7%	92.6%	95.2%
lytA hylA	93.1%	87.2%	100.0%	100.0%
piaA hylA	92.1%	90.0%	92.1%	94.9%
lytA piaA hylA	92.1%	86.2%	92.1%	94.6%

lytA

	Pneumococcal	Non-pneumococcal
Test positive	150	2
Test negative	0	126

$PPV = \frac{150}{150+2} \times 100\%$ PPV= 98.7%

$NPV = \frac{126}{126+0} \times 100\%$ NPV= 100.%

piaA

	Pneumococcal	Non-pneumococcal
Test positive	140	1
Test negative	10	127

$PPV = \frac{140}{140+1} \times 100\%$ PPV= 99.3%

$NPV = \frac{127}{127+10} \times 100\%$ NPV= 93.4%

hylA

	Pneumococcal	Non-pneumococcal
Test positive	150	11
Test negative	0	117

$PPV = \frac{150}{140+11} \times 100\%$ PPV= 93.1%

$NPV = \frac{117}{117+0} \times 100\%$ NPV= 100%

lytA piaA

	Pneumococcal	Non-pneumococcal
Test positive	140	3
Test negative	10	125

$PPV = \frac{140}{140+3} \times 100\%$ PPV= 97.9%

$NPV = \frac{125}{125+10} \times 100\%$ NPV= 92.6%

lytA hylA

	Pneumococcal	Non-pneumococcal
Test positive	150	13
Test negative	0	115

$PPV = \frac{150}{150+13} \times 100\%$ PPV= 93.1%

$NPV = \frac{115}{115+0} \times 100\%$ NPV= 100%

lytA piaA hylA

	Pneumococcal	Non-pneumococcal
Test positive	140	12
Test negative	10	116

$PPV = \frac{140}{140+10} \times 100\%$ PPV= 92.1%

$NPV = \frac{116}{116+10} \times 100\%$ NPV= 92.1%

Supplementary Figure 1.PPV and NPV calculation for single and combined qPCR assays.