TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. D OSA - TOM. 961

MEDICA - ODONTOLOGICA

USE OF PYROSEQUENCING TO IDENTIFY STREPTOCOCCI AND TO DETECT MUTATIONS CAUSING ANTIMICROBIAL RESISTANCE

by

Marjo Haanperä-Heikkinen

TURUN YLIOPISTO
UNIVERSITY OF TURKU
Turku 2011

From the Department of Medical Microbiology and Immunology, University of Turku, Turku, Finland and the Antimicrobial Resistance Unit, National Institute for Health and Welfare (former National Public Health Institute), Turku, Finland.

Supervised by

Docent Jari Jalava, MSc, PhD
Antimicrobial Resistance Unit
National Institute for Health and Welfare
Turku, Finland

and

Professor Pentti Huovinen, MD, PhD
Department of Medical Microbiology and Immunology
University of Turku
Turku, Finland

Reviewed by

Docent Risto Vuento, MD, PhD
Centre for Laboratory Medicine
Tampere University Hospital
Tampere, Finland.

and

Baback Gharizadeh, MSc, PhD
Stanford Genome Technology Center
Stanford University
Palo Alto, CA, U.S.A.

Opponent

Professor Simo Nikkari, MD, PhD
Centres for Biothreat Preparedness and Military Medicine
Helsinki, Finland

ISBN 978-951-29-4598-6 (PRINT)
ISBN 978-951-29-4599-3 (PDF)
ISSN 0355-9483
Painosalama Oy – Turku, Finland 2011

To my family

ABSTRACT

Marjo Haanperä-Heikkinen
Use of pyrosequencing to identify streptococci and to detect mutations causing antimicrobial resistance
The Department of Medical Microbiology and Immunology, University of Turku, Finland and the Antimicrobial Resistance Unit, National Institute for Health and Welfare, Turku, Finland
Annales Universitatis Turkuensis
Painosalama Oy - Turku, Finland 2011

Rapid identification and resistance determination of pathogens in clinical specimens is vital for accurate treatment and monitoring of infectious diseases. Antimicrobial drug resistance is increasing globally and healthcare settings are facing this cost-intensive and even life-threatening problem. The incidence of resistant pathogens in Finland has remained relatively steady and manageable at least for the time being.

DNA sequencing is the gold standard method for genotyping, mutation analysis, and identification of bacteria. Due to significant cost decrease in recent years, this technique is available to many research and clinical laboratories. Pyrosequencing technique, a rapid real-time DNA sequencing method especially suitable for analyzing fairly short stretches of DNA, was used in this study. Due to its robustness and versatility, pyrosequencing was applied in this study for identification of streptococci and detection of certain mutations causing antimicrobial resistance in different bacteria.

Certain streptococcal species such as *S. pneumoniae* and *S. pyogenes* are significantly important clinical pathogens. *S. pneumoniae* causes e.g. pneumonia and otitis media and is one of the most important community-acquired pathogens. *S. pyogenes*, also known as group A streptococcus, causes e.g. angina and erysipelas. In contrast, the so-called alpha-haemolytic streptococci, such as *S. mitis* and *S. oralis,* belong to the normal microbiota, which are regarded to be non-pathogenic and are nearly impossible to identify by phenotypic methods. In this thesis, a pyrosequencing method was developed for identification of streptococcal species based on the 16S rRNA sequences. Almost all streptococcal species could be differentiated from one another by the developed method, including *S. pneumoniae* from its close relatives *S. mitis* and *S. oralis*.

New resistance genes and their variants are constantly discovered and reported. In this study, new methods for detecting certain mutations causing macrolide resistance or extended spectrum beta-lactamase (ESBL) phenotype were developed. These resistance detection approaches are not only suitable for surveillance of mechanisms causing antimicrobial resistance but also for routine analysis of clinical samples particularly in epidemic settings.

In conclusion, pyrosequencing was found to be an accurate, versatile, cost-effective, and rapid DNA sequencing method that is especially suitable for mutation analysis of short DNA fragments and identification of certain bacteria.

Keywords: pyrosequencing, streptococci, antimicrobial resistance, 16S rRNA, 23S rRNA, macrolides, ESBL, SHV

TIIVISTELMÄ

Marjo Haanperä-Heikkinen
Pyrosekvensoinnin käyttö streptokokkien tunnistamiseen ja mikrobilääkeresistenssiä aiheuttavien mutaatioiden havaitsemiseen
Lääketieteellinen mikrobiologia ja immunologia, Turun yliopisto ja Mikrobilääkeresistenssiyksikkö, Terveyden ja hyvinvoinnin laitos, Turku, Suomi
Annales Universitatis Turkuensis
Painosalama Oy – Turku, Suomi 2011

Kliinisen näytteen sisältämien patogeenien nopea tunnistaminen ja resistenssiprofiilin määrittäminen on välttämätöntä infektiotautien tehokkaassa hoidossa ja seurannassa. Bakteerien lisääntyvä antibiootti- eli mikrobilääkeresistenssi on merkittävä uhka terveydenhuollolle kaikkialla maailmassa. Suomessa mikrobilääkeresistenssitilanne on pysynyt suhteellisen vakaana ja hallittavissa olevana ainakin toistaiseksi.

DNA:n sekvensointi on bakteerien tunnistamisen, genotyypityksen ja mutaatioanalyysin yleisesti hyväksytty vertailumenetelmä. Sekvensointitekniikan hinnan merkittävä aleneminen on tuonut sekvensoinnin useimpien kliinisten ja tutkimuslaboratorioiden ulottuville. Tässä työssä käytettiin pyrosekvensointitekniikkaa, nopeaa reaaliaikaista DNA:n sekvensointimenetelmää, joka soveltuu erityisesti lyhyiden DNA-jaksojen analysoimiseen. Luotettavaa ja monikäyttöistä pyrosekvensointitekniikkaa sovellettiin tässä työssä streptokokkien tunnistamiseen ja tiettyjen antibioottiresistenssiä aiheuttavien geenimutaatioiden havaitsemiseen eri bakteereista.

Tietyt streptokokit, kuten *S. pneumoniae* ja *S. pyogenes,* ovat merkittäviä taudinaiheuttajia. *S. pneumoniae* aiheuttaa mm. keuhkokuumetta ja korvatulehduksia ja on avohoidon tärkeimpiä taudinaiheuttajia. *S. pyogenes* eli A-streptokokki aiheuttaa mm. angiinaa ja ruusua. Toisaalta ns. alfahemolyyttisiä streptokokkeja, kuten *S. mitis* ja *S. oralis,* kuuluu myös apatogeenisenä pidettyyn normaalimikrobistoon. Alfahemolyyttisten streptokokkien tunnistaminen fenotyyppisin menetelmin on lähes mahdotonta. Tässä työssä kehitetyllä 16S rRNA -sekvensseihin perustuvalla pyrosekvensointimenetelmällä pystyttiin erottamaan lähes kaikki streptokokkilajit toisistaan, mukaan lukien *S. pneumoniae* lähisukulaisistaan *S. mitis* - ja *S. oralis* -kannoista.

Mikrobeista löydetään yhä uusia resistenssigeenejä ja niiden variantteja. Tässä työssä kehitettiin menetelmät tiettyjen makrolidiresistenssiä tai laajakirjoista beetalaktamaasifenotyyppiä (ESBL) aiheuttavien mutaatioiden havaitsemiseksi. Tässä työssä kehitettyjen kaltaiset resistenssimutaatioita havaitsevat menetelmät soveltuvat erityisesti mikrobilääkeresistenssin mekanismien seurantaan, mutta myös potilasnäytteiden rutiinianalytiikkaan erityisesti epidemiatilanteissa.

Tässä työssä pyrosekvensoinnin havaittiin olevan tarkka, joustava, edullinen ja nopea sekvensointimenetelmä, joka soveltuu erityisesti mutaatioiden analysoimiseen ja tiettyjen bakteerien tunnistamiseen.

Asiasanat: pyrosekvensointi, streptokokit, mikrobilääkeresistenssi, 16S rRNA, 23S rRNA, makrolidit, ESBL, SHV

CONTENTS

ABBREVIATIONS

16S rRNA	16S ribosomal RNA
ADP	Adenosine diphosphate
ampC	A beta-lactamase that hydrolyses broad and extended-spectrum cephalosporins
ANI	Average nucleotide identity
APS	Adenosine phosphosulfate
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
Bst	*Bacillus stearothermophilus*
CCD	Charge-coupled device
CCUG	Culture Collection, University of Göteborg, Gothenburg, Sweden
dATP	Deoxyadenosine triphosphate
dATP-α-S	2'-deoxyadenosine-5'-O'- (1-thiotriphosphate), dATP analogue
DDH	DNA-DNA hybridization
ddl	D-alanine:D-alanine ligase gene
ddNTP	Dideoxynucleoside triphosphate
dNMP	Deoxynucleoside monophosphate
dNTP	Deoxynucleoside triphosphate, in the text often also "nucleotide"
DSM	Deutsche Sammlung von Mikroorganismen GmbH, Braunschweig, Germany
emPCR	emulsion PCR
Erm	Erythromycin ribosomal methylase
ESBL	Extended spectrum β-lactamase
GAS	Group A streptococcus
GBS	Group B streptococcus
groESL	Genes encoding 10-kDa (GroES) and 60-kDa (GroEL) chaperons
GenBank	The genetic sequence database of National Institutes of Health (Bethesda, MD, USA) an annotated collection of all publicly available DNA sequences
HGT	Horizontal gene transfer
HPLC	High-performance liquid chromatography
IE	Infective endocarditis
LMG	Laboratorium voor Mikrobiologie, Universiteit Ghent, Ghent, Belgium
lytA	Cell wall autolysin gene
MALDI-TOF	Matrix-assisted laser desorption ionization-time-of-flight
Mef	Macrolide efflux pump
MIC	Minimal inhibitory concentration

ML	Macrolide-lincosamide resistance phenotype
MLS_B	Macrolide-lincosamide-streptogramin B resistance phenotype
MLSA	Multi locus sequence analysis
MS	Mass spectrometry
NCBI	National Center for Biotechnology Information
OXA	Beta-lactamase that hydrolyses oxacillin and cloxacillin
PBP	Penicillin binding protein
PCR	Polymerase chain reaction
Ply	Pneumolysin
Pi	(Inorganic) phosphate
PPi	Pyrophosphate
PTC	Peptidyl transferase center
RDP	Ribosomal database project
rpoB	Gene encoding the β-subunit of the RNA polymerase
SHV	Sulfhydryl variable, a beta-lactamase
SNP	Single Nucleotide polymorphism
sp.	Species
ssp.	Subspecies
SSB	Single-strand binding protein
SSCP	Single-strand conformational polymorphism
TEM	A beta-lactamase named after Temoneira
Tm	Melting temperature
tuf	Gene encoding the elongation factor Tu (thermo unstable)
VGS	Viridans group streptococci
VP	Voges-Proskauer test

LIST OF ORIGINAL COMMUNICATIONS

This thesis is based on the following publications, which are referred to in the text by the Roman numerals I-IV. The original communications are reproduced with the kind permission from the copyright holders, The British Society for Antimicrobial Chemotherapy (I) or the American Society for Microbiology (II- IV).

I Seppälä, H., Haanperä, M., Al-Juhaish, M., Järvinen, H., Jalava, J. and Huovinen, P. Antimicrobial susceptibility patterns and macrolide resistance genes of viridans group streptococci from normal flora. Journal of Antimicrobial Chemotherapy 2003; 52:636-644.

II Haanperä, M., Jalava, J., Huovinen, P., Meurman, O. and Rantakokko-Jalava, K. Identification of alpha-hemolytic streptococci by pyrosequencing the 16S rRNA gene and by use of Vitek 2. Journal of Clinical Microbiology. 2007; 45: 762-770.

III Haanperä, M., Huovinen, P. and Jalava, J. Detection and quantification of macrolide resistance mutations at positions 2058 and 2059 of the 23S rRNA gene by pyrosequencing. Antimicrobial Agents and Chemotherapy. 2005; 49:457-460

IV Haanperä, M.*, Nyberg S.D.*, Huovinen, P. and Jalava, J. Typing of SHV ESBL β-lactamases by pyrosequencing in *Klebsiella pneumoniae* strains with chromosomal SHV β-lactamase. Antimicrobial Agents and Chemotherapy. 2008; 52:2632-2635.

* These two authors contributed equally to the work.

1 INTRODUCTION

The most important tasks of a clinical microbiology laboratory are identification of the infectious agents found in a specimen and determination of their antimicrobial resistance profile. At present, bacterial diagnostics in routine laboratories mainly relies on culturing and semi-automatic platforms running biochemical tests. However, use of DNA sequencing has become more common mainly because 16S rRNA sequencing has been shown to be useful in identification of bacterial isolates and in describing new bacterial species (Woo *et al.* 2008) and also because the cost of sequencing has decreased significantly.

Antimicrobial resistance is globally an increasing problem. Patients infected with resistant pathogens are more likely to require longer and more expensive hospital stay and they are more likely to die of the infection. Particularly, patients infected with MDR (multidrug resistant) strains are difficult to treat, as treatment options may be few, often less effective, more toxic, and more expensive.

Macrolide resistance has been common in *S. pyogenes* (Seppälä *et al.* 1992; Seppälä *et al.* 1993), but also the incidence of macrolide resistant *S. pneumoniae* isolates has increased during the last fifteen years (Bergman *et al.* 2006; Pihlajamäki *et al.* 2002a). During the past ten years, enterobacteria expressing extended spectrum beta-lactamases (ESBLs) causing resistance to several beta-lactam antibiotics, have spread rapidly and caused serious difficulties in the treatment of patients infected by these bacteria. Currently, several hundred different ESBLs have been described, and detection of bacterial isolates producing ESBL is complicated due to the wide variety of the enzymes and their resistance profiles (Jacoby and Bush 2010; Pitout and Laupland 2008).

Streptococci are clinically important pathogens and at the same time members of the normal microbiota of human and animals. Streptococci are divided into alpha- and beta-haemolytic streptococci. Beta-haemolytic streptococci are generally considered as pathogens and easily identified in the clinical laboratory. In contrast, the alpha-haemolytic streptococci are difficult to identify, and they are also more often found in the normal microbiota.

Pyrosequencing technique is based on the detection of pyrophosphate released during DNA synthesis. Pyrosequencing technology was developed by Pål Nyrén and his colleagues (Ronaghi *et al.* 1998). The aim of this thesis project was to develop accurate, robust, simple, and cost-effective methods for the identification of streptococci, macrolide resistance mutations, and different SHV β-lactamase types. Pyrosequencing was the method of choice as it was a novel innovative sequencing technique at the beginning of this thesis project and our institute had the privilege of purchasing the requisite. This study did not aim to further develop and improve the pyrosequencing technique.

2 REVIEW OF THE LITERATURE

2.1 Pyrosequencing technique

Pyrosequencing is a PCR-based real-time DNA sequencing method, also called sequencing by synthesis. Pyrosequencing is an non-electrophoretic method that is based on enzymatic detection of the pyrophosphate that is released when DNA polymerase incorporates a nucleotide into an elongating DNA strand (Ronaghi *et al.* 1998).

Sequencing by synthesis was first described in a patent in 1985 (Melamede 1989). The patent of Melamede described a method where nucleotides were added sequentially to a primed template and the incorporation of a nucleotide was detected by decrease of the nucleotide concentration in the effluent. However, this method was not sensitive enough for standard PCR-amplified samples (Nyrén 2007). In 1987, Pål Nyrén followed the activity of DNA polymerase by enzymatically detecting the pyrophosphate released in DNA polymerization using ATP-sulfurylase and firefly luciferase (Nyrén 1987). The same enzymes are used in pyrosequencing reactions today.

The pyrosequencing method including the nucleotide-degrading enzyme apyrase was published in 1998. Addition of apyrase to the reaction made the washing step between nucleotide additions unnecessary (Ronaghi *et al.* 1998). The four-enzyme pyrosequencing reaction and instrument for it was commercialized by Pyrosequencing AB (Uppsala, Sweden), Biotage AB since 2003. Biotage was acquired by Qiagen (Hilden, Germany) in 2008. In addition, pyrosequencing technique was licensed to 454 Life Sciences (currently owned by Roche, Basel, Switzerland) in 2003 to be used in whole genome applications. Accordingly, 454 Life Sciences has developed an array-based high-throughput pyrosequencing platform that is extensively used for large projects, such as genome sequencing (Bonnal *et al.* 2010; Margeridon-Thermet *et al.* 2009; Quince *et al.* 2009). This technique is discussed in chapter 2.1.7 of this thesis.

2.1.1 Template preparation

The PCR product to be used as a template in pyrosequencing should be biotinylated and relatively short, preferably no longer than 300 bp to prevent the single-stranded PCR product from forming strong secondary structures during pyrosequencing reaction.

During template preparation, the biotin-labelled PCR product is attached to streptavidin-coated beads. The beads are sepharose or magnetic beads and with help of these beads, the PCR product is rendered single-stranded using magnetic separation techniques or the commercial Vacuum Workstation (Qiagen). Using the Vacuum Workstation, the beads are denatured using 0.2 M NaOH solution and thereafter washed with a buffer to neutralize the pH. The beads with the single-stranded template are deposited into the pyrosequencing reaction well containing the sequencing primer where the primer is annealed to the template.

Additional primer wash (Gharizadeh *et al.* 2006; Gharizadeh *et al.* 2003) after annealing the primers to the template reduces the problems with primer-dimers which may be a problem especially when using multiple sequencing primers. However, this additional step requires more PCR product to be applied to the sequencing reaction.

2.1.2 Pyrosequencing reaction

The pyrosequencing enzymes, Klenow fragment of DNA polymerase 1, ATP sulfurylase, firefly luciferase, and apyrase as well as their substrates adenosine phosphosulfate (APS) and D-luciferin are added to the reaction well containing the primed template. The sequencing reaction is started by dNTP addition to the reaction mixture. If the given nucleotide is complementary, the DNA polymerase incorporates the nucleotide into the elongating strand and pyrophosphate (PPi) is released (Ronaghi *et al.* 1998):

$$(DNA)_n + dNTP \rightarrow (DNA)_{n+1} + PPi$$

The PPi is further processed by ATP sulfurylase that produces ATP in equimolar amount to the PPi in presence of the substrate APS:

$$PPi + APS \rightarrow ATP + SO_4^{2-}$$

Finally, firefly luciferase consumes the ATP in producing light from the substrate luciferin. Light is generated in proportion to the amount of pyrophosphate released, thus the more nucleotides are incorporated the more light is produced:

Luciferase + D-luciferin + ATP $\rightarrow$ Luciferase – luciferin – AMP + PPi

Luciferase – luciferin – AMP + O_2 $\rightarrow$ Luciferase + oxyluciferin + AMP + CO_2 + light

In the commercial, four-enzyme system, unincorporated nucleotides and ATP are degraded by potato apyrase, a nucleotide degrading enzyme, before the next nucleotide is added (Ronaghi 2001; Ronaghi *et al.* 1998):

ATP $\rightarrow$ ADP + Pi $\rightarrow$ AMP+Pi

dNTP $\rightarrow$ dNMP + 2 Pi

The light released by the luciferase reaction is detected by a charge-coupled device (CCD) camera and presented as a signal peak illustration, called pyrogram (Figure 1). In the pyrogram, the dispensed nucleotides are presented on the x-axis and the light detected after a nucleotide addition is presented as a peak on the y-axis. The peak heights of the pyrogram are proportional to the number of nucleotides incorporated into the template strand.

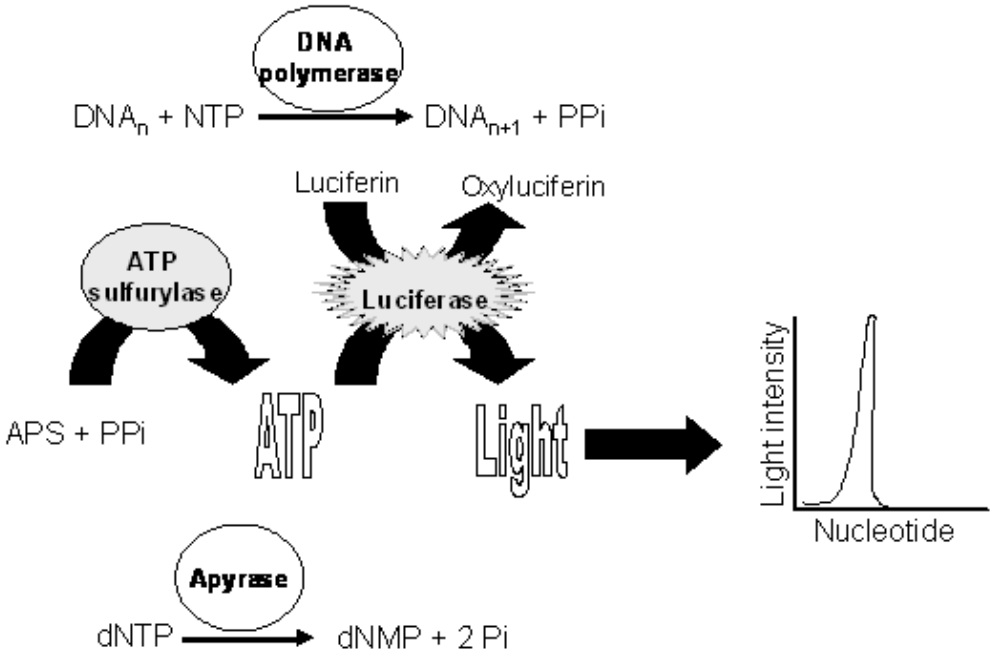


Figure 1. The principle of pyrosequencing.

Incorporation of a single base by the Klenow fragment happens in less than 0.5 s and light production by luciferase happens in less than 0.2 s. The rate limiting factors in pyrosequencing are the conversion of PPi to ATP by ATP sulfurylase (around 1.5 sec) and the degradation of nucleotides by apyrase. Consequently, the whole cycle from nucleotide dispensation to peak formation and degradation of the unused nucleotides takes place in around 5 sec (Ronaghi 2001; Ronaghi *et al.* 1996). Currently nucleotides are given to the reaction with 65 seconds interval in the instruments of Qiagen.

2.1.3 Enzymes and dATP analogue

The performance of all four enzymes is vital for the pyrosequencing technology, and they all have properties that affect the performance of pyrosequencing.

DNA polymerase

The DNA polymerases (E.C. 2.7.7.7) catalyze DNA polymerization during cell replication and DNA repair and are crucial for all living cells. There are several requirements for the DNA polymerase used in pyrosequencing: deficiency of the 3'-5' exonuclease, good fidelity, reasonable speed of nucleotide incorporation, ability to incorporate dATP analogues, and ability to incorporate nucleotides reliably through homopolymeric regions. The polymerase usually used in pyrosequencing is the modified Klenow fragment of *Escherichia coli* polymerase I: two amino acids of the Klenow fragment have been mutated to inactivate the 3'-5' exonuclease activity of the polymerase (Gharizadeh *et al.* 2004). Deficiency of the 3'-5' exonuclease activity is

important in order to avoid primer degradation and non-synchronized extension of the templates. Although the enzyme is proofreading-deficient, other mechanisms make the incorporation of the correct base more probable than a wrong one. Firstly, the complementary base is incorporated into the template more tightly than a non-complementary base. Secondly, the kinetics of the incorporation of a non-complementary base is much slower than the kinetics of the complementary base, which allows apyrase to degrade the non-complementary nucleotides before possible misincorporations into the growing strand (Ahmadian *et al.* 2006; Ahmadian *et al.* 2001).

Exonuclease deficient T7 DNA polymerase, Sequenase has also been tested in pyrosequencing. Sequenase has several advantages in pyrosequencing when compared to the Klenow fragment: Sequenase incorporates dATP-α-S nucleotides to homopolymeric T stretches more efficiently, it does not accept mismatches in its substrate and therefore, unspecifically bound sequencing primers are not extended. In addition, Sequenase does not extend primer-dimers or loop structures unless they have at least 15 base pairs long stem (Gharizadeh *et al.* 2004). The large subunit of the *Bst* polymerase isolated from *Bacillus stearothermophilus* is used in massively parallel pyrosequencing (Margulies *et al.* 2005).

ATP sulfurylase

ATP sulfurylase (E.C. 2.7.7.4) catalyzes the production of ATP from PPi in pyrosequencing. *In vivo,* ATP sulfurylase activates sulphur by producing APS from ATP and SO_4^{2-}, and the APS is used for synthesis of different sulphur containing molecules. However, the reaction kinetics is very unfavourable for APS production but the removal of APS and PPi by cellular enzymes drives the reaction to the right. Thus, the reaction kinetics is favourable for the reaction needed in pyrosequencing: production of ATP from APS and PPi. The ATP sulfurylase used in pyrosequencing is a recombinant enzyme from *Saccharomyces cerevisiae* (Karamohamed *et al.* 1999).

Luciferase

Luciferase (E.C. 1.13.12.7) catalyses the production of light from ATP in pyrosequencing. The luciferase used in pyrosequencing is the most studied and the only commercially available recombinant luciferase originating from North American firefly *Photinus pyralis*. However, a more thermostable luciferase would allow pyrosequencing to be performed at a higher temperature (Ahmadian *et al.* 2006). A higher sequencing temperature would accelerate the sequencing reactions and reduce the background signal by diminishing the formation of primer dimers and strong secondary structures in the template (Eriksson *et al.* 2004).

Apyrase

The apyrase (E.C. 3.6.1.5.) used in pyrosequencing derives from potato, *Solanum tuberosum*, pimpernel variety (Ahmadian *et al.* 2006). The use of apyrase enabled liquid-phase pyrosequencing reaction since apyrase degrades the unincorporated dNTPs and ATP from the reaction between different nucleotide additions (as opposed to solid-phase pyrosequencing). If apyrase is not used, pyrosequencing templates are washed and new enzymes and substrates are added between nucleotide dispensations (Langaee and Ronaghi 2005).

Single-strand binding protein

Single-strand binding protein (SSB) improves the read length of pyrosequencing by releasing secondary structures that hinder the action of DNA polymerase. The SSB used in pyrosequencing is isolated from *E. coli* (Ronaghi 2000). SSB is currently included in all commercial pyrosequencing kits.

dATP analogue

The natural dATP is a weak substrate of luciferase (Maxam and Gilbert 1977; Ronaghi *et al.* 1996). Therefore, use of dATP would lead to significant background signals in pyrosequencing because the amount of nucleotides given to the reaction is approximately 10 times higher than the amount of template (Zhou *et al.* 2005). Consequently, the nucleotide analogue 2'-Deoxyadenosine-5'-O'- (1-thiotriphosphate), (alpha-thio A, dATP-α-S) is used instead of dATP in pyrosequencing; dATP-α-S is not a substrate of luciferase whereas the Klenow fragment can use it as a substrate (Ronaghi *et al.* 1996). Due to the different structure, the A peaks are approximately 5-15% higher in pyrosequencing than the peaks of other nucleotides.

The dATP analogue has caused problems in pyrosequencing because it had an inhibitory effect on both the Klenow fragment and apyrase. As a consequence, homopolymeric T regions on the template were especially problematic in pyrosequencing (Ronaghi *et al.* 1999). Later it was found that only the Sp isomer of dATP-α-S was utilized by the Klenow fragment, and the Rp isomer is an inhibitor (Burgers and Eckstein 1979; Gharizadeh *et al.* 2002). When the pure Sp isomer was used, the alpha-thio A concentration could be halved and sequences of 50-100 bases were obtainable (Gharizadeh *et al.* 2002).

dATP-α-S has an inhibitory effect also on apyrase, and apyrase activity was continuously decreasing during pyrosequencing reaction when both isoforms of the dATP-α-S were added to the reaction. The apyrase inhibition is however probably primarily due to the accumulation of the end products, mainly nucleoside diphosphates. Consequently, apyrase inhibition is not dependent on the stereoisomer, but the difference in inhibition is due to the lower concentration of the Sp isomer needed in the sequencing reaction (Gharizadeh *et al.* 2002).

2.1.4 Dispensation order and pyrogram

The CCD camera in pyrosequencing instrument detects the light produced by the luciferase-luciferin reaction and records and presents the light emission as a sequence signal peaks called pyrogram. In the pyrogram, the detected light intensity is presented on the y-axis and the order of nucleotide dispensations are presented on the x-axis. The light emission is quantitative and proportional to the number of nucleotides incorporated; the more nucleotides incorporated, the more light is emitted, and the higher pyrosequencing peak appears to the pyrogram.

The dispensation order, or the order in which the nucleotides are added to the pyrosequencing reaction, can be freely selected. This property enables to design sequence-specific dispensation orders, which is very useful when known nucleotide sequences as single nucleotide polymorphisms (SNPs) are analyzed (Figure 2). However, simple cyclic dispensation order can also be used. Using calculated, predetermined dispensation order in pyrosequencing enables longer reads as less inhibitory end products are accumulated in the reaction.

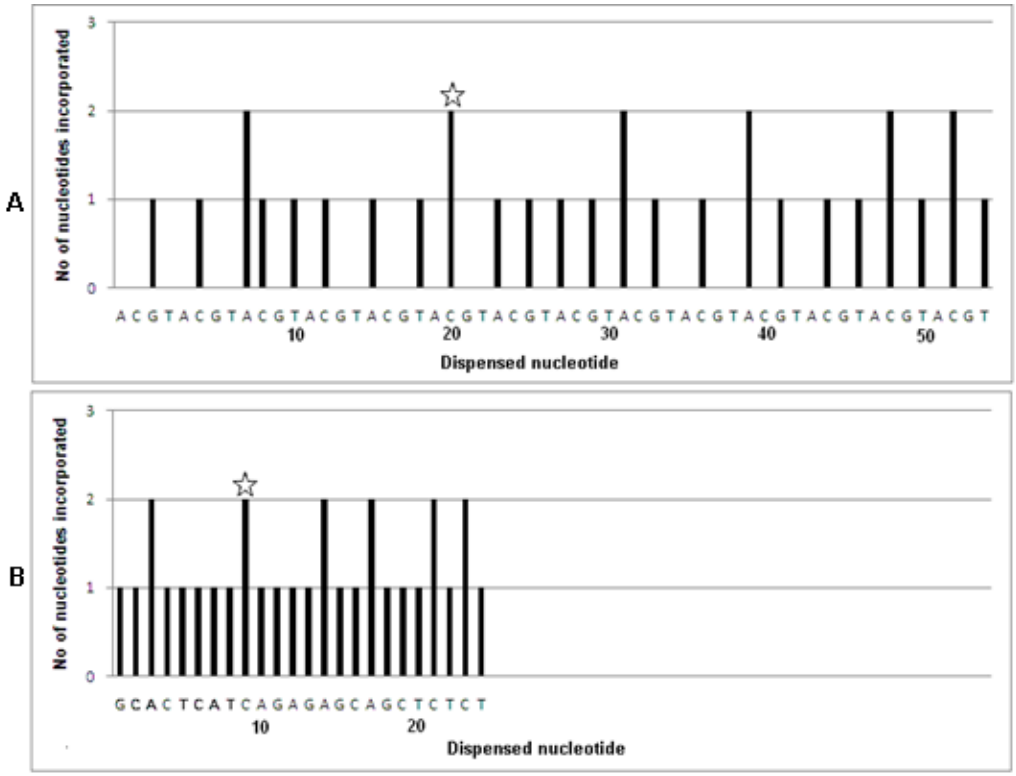


Figure 2. Theoretical pyrograms of the 30-bp sequence of the *S. pneumoniae* v1 region obtained using cyclic (A) or a predetermined, *S. pneumoniae* –specific dispensation order (B). The star in the figure illustrates a hypothetical mutation site that is reached after twenty dispensations using the cyclic dispensation order whereas nine dispensations suffice when using the predetermined dispensation order.

Pyrosequencing may also be performed in a multiplex fashion (Pourmand *et al.* 2002), using more than one sequencing primer that anneal to different sites. The sequences are determined from the pyrosequencing peak patterns that in principle are the sum-curves of the known sequences (Figure 3). With the current pyrosequencing software it is possible to analyse three SNP targets per reaction, with more sophisticated software probably more SNPs per reaction can be analysed (Langaee and Ronaghi 2005).

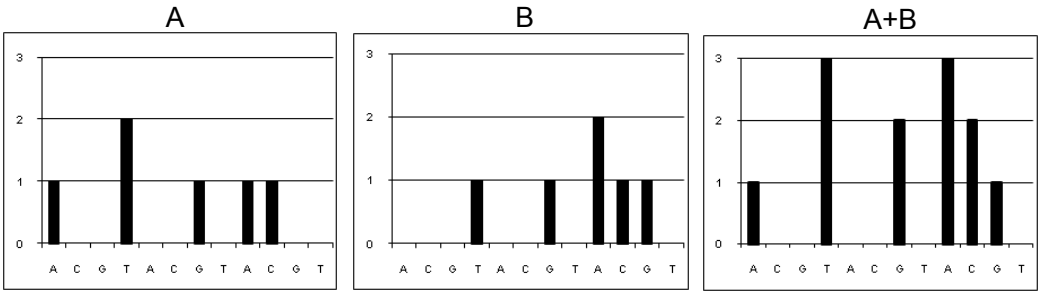


Figure 3. Multiplex pyrosequencing, the same (cyclic) dispensation order is used in all figures. A) The sequence in the pyrogram is ATTCAC and B) the sequence is TGAACG. When these sequencing reactions are combined, the sum curve of A and B is generated (A+B).

2.1.5 Problems inherent to pyrosequencing

The pyrosequencing enzyme system has some properties that may lead to difficulties during pyrosequencing. However, these known errors do not affect the generated sequence results. The main factors limiting read-length of pyrosequencing are appearance of background sequence and minus and/or plus frame shifts. The amount of background sequence is proportional to length of the PCR product: the longer the product, the more background signal is generated (Ahmadian *et al.* 2006).

Minus shift is generated when a complementary nucleotide is not incorporated to all possible templates by the DNA polymerase and nucleotides are incorporated to them at later dispensations of the same nucleotide (Svantesson *et al.* 2004). This leads to small peaks after the correct dispensation and may later lead to the templates being sequenced asynchronously, which in turn leads to spurious pyrosequencing results, which are hard to interpret.

The appearance of background signal before the real peak is called plus shift. It is due to incomplete degradation of nucleotides by apyrase, which leads to intact nucleotides being present at the next dispensation, leading to false sequence signals that might interfere with result interpretations in longer reads (Ahmadian *et al.* 2006).

The decrease of signals towards the end of long read sequencing runs is caused by the inhibition of enzymes by the end by-products, mainly nucleoside monophosphates (dNMPs) (Svantesson *et al.* 2004), and the dilution effect due to iterative nucleotide additions (0.2 µl/ dispensation). The dilution effect can mainly be seen as broadening of peaks and the lowering of signals at later dispensations. Dilution effect is, however, more significant in so-called long-read sequencing where sequencing more than 100 bases requires approximately 150 nucleotide dispensations (Gharizadeh *et al.* 2002).

Sequencing homopolymeric DNA stretches is difficult for pyrosequencing technology, as the light generated from more than three nucleotides a dispensation may not increase linearly particularly when it involves alpha-thio A incorporations. This is mainly due to the longer time needed for nucleotide incorporation by DNA polymerase and simultaneous degradation of nucleotides by apyrase (Pourmand *et al.* 2002). However, adjacent sequence peaks and frameshifts may be used to determine the exact number of incorporated nucleotides.

Pyrosequencing has also an inherent property to unsynchronized extension after a polymorphic site. This can be avoided using a predetermined, non-cyclic dispensation in order to obtain a synchronized reaction after the polymorphic site (Ahmadian *et al.* 2006). The use of programmed nucleotide dispensation order also increases the read-length because most nucleotide additions lead to extension and less end-products are accumulated in the reaction (Garcia *et al.* 2000).

2.1.6 Applications of pyrosequencing

As pyrosequencing can be used for sequencing short stretches of DNA, it can be used in various applications. It has been found useful for example in microbial typing and detection of mutations causing resistance to antimicrobial agents in viruses (Lindström *et al.* 2004; O'Meara *et al.* 2001) and bacteria (Moder *et al.* 2007; Naas *et al.* 2007; Naas *et al.* 2006; Poirel *et al.* 2006). Pyrosequencing has also been widely used for determining SNPs and methylation analysis in genomic DNA (Baba *et al.* 2010; Fakhrai-Rad *et al.* 2002).

Sequencing

Pyrosequencing can be used for actual sequence determination, provided that the sequence to be determined by a single pyrosequencing run is fairly short as the maximum read length is normally approximately 60 bases (Mashayekhi and Ronaghi 2007; Ronaghi 2000). In addition, the appearance of nonsynchronized sequencing data after a challenging polymorphic region limits the use of pyrosequencing mainly to re-sequencing. However, the sequence of exons 5-8 of the p53 gene has been determined by pyrosequencing, the sequence was assembled from short overlapping sequences determined by pyrosequencing (Garcia *et al.* 2000). Furthermore, the large-scale pyrosequencing technique has been used in several genome sequencing projects (Monger *et al.*; Wheeler *et al.* 2008).

Detection of SNPs or heterogeneous sequences

The inherent property of a pyrosequencing reaction to proceed asynchronously if the templates are not identical can be exploited in detecting SNPs or heterogeneous sequences (Lindström *et al.* 2004) as the dispensation order may be designed so that the different haplotypes: wild type, mutant and heterozygous create a distinct pyrosequencing peak pattern. In fact, the commercial pyrosequencing technology (Qiagen) has been developed mainly for determination of mutations, mostly the SNPs present in the human genome. Pyrosequencing is suitable for detection of SNPs since they are short nucleotide stretches, which can easily be determined by pyrosequencing. Furthermore, as pyrosequencing is a sequencing method, also the nucleotides surrounding the SNP of interest are sequenced verifying the correct positioning of the assay and novel mutations can be detected.

Pyrosequencing is useful also for molecular haplotyping, determination of whether a given mutation is present on both chromosomes or only on other chromosomal copy (Ahmadian *et al.* 2000) or whether given sequences are located *cis* or *trans* in the chromosomes (Kruckeberg and Thibodeau 2004). Pyrosequencing analysis may also be performed with pooled samples to determine the proportion of a given sequence type in a pooled population. The pooling method is useful for screening of proportion of a given SNP in DNA pools from affected or non-affected individuals before implementing more detailed methods (Doostzadeh *et al.* 2008; Gruber *et al.* 2002).

In addition to SNPs, pyrosequencing is suitable for determination of other kinds of heterogeneous sequences, for example a sequence of a virus or a bacterium that carries

a resistance mutation against a certain drug (Lindström *et al.* 2004; Sinclair *et al.* 2003). In a study for determining drug resistance of hepatitis B virus, pyrosequencing could detect a heterogeneous sequence that was present in 20% of the templates (Lindström *et al.* 2004), and in linezolid resistant enterococci, one mutant copy among six copies was detectable (Sinclair *et al.* 2003).

Methylation detection

Pyrosequencing has also been applied to the detection of methylated sequences that are important for example in the development of cancer because methylation affects the gene expression profiles and changes in DNA methylation often occur early in carcinogenesis. During cancer development, two kinds of methylation changes have been detected: either the whole genome is hypomethylated or cytosine residues in promoter-associated CpG islands are over- or undermethylated (Shames *et al.* 2007).

Prior to pyrosequencing methylation analysis, the sample DNA is treated by bisulphite, which converts unmethylated cytosines to urasils whereas methylated cytosines remain unchanged. When the bisulphite-treated DNA is amplified by PCR, the urasils are replaced by thymidines and the cytosines remain cytosines. In pyrosequencing analysis of the PCR product, the proportion of C peak to T peak indicates the degree of methylation at a given site (England and Pettersson 2005).

Identification of bacteria

Identification of bacteria using pyrosequencing utilizes DNA targets that have conserved and variable regions so that the primers can be designed to the conserved areas and pyrosequencing is performed on the variable regions (Grahn *et al.* 2003; Innings *et al.* 2005; Jonasson *et al.* 2002; Jordan *et al.* 2005; Monstein *et al.* 2001; Ronaghi and Elahi 2002; Tuohy *et al.* 2005; Unnerstad *et al.* 2001). First pyrosequencing assay for the identification of bacteria was published by Jonasson *et al* in 2002. The method was based on the variable regions v1 and v3 of the 16S rRNA that have also been used for the identification of contaminating bacterial DNA in PCR reagents and for detection and subtyping of *Helicobacter pylori* isolates (Grahn *et al.* 2003; Monstein *et al.* 2001).

Moreover, a method using a universal 16S rRNA PCR for template preparation and multiple pyrosequencing primers for detection of a certain bacterial groups was developed. In this method, multiple semi-conservative sequencing primers were designed to give the most informative sequence in different bacterial species. During pyrosequencing, only one of the primers hybridises to the target sequence of a given species (Gharizadeh *et al.* 2003). However, the sequences of ribosomal genes are sometimes identical in closely related bacteria and analysis of other, more variable house-keeping genes or genes specific to the target organism have been introduced. For example, *Bacillus anthracis* has been differentiated from other *Bacillus* species by a pyrosequencing assay based on determination of four SNPs of the RNA polymerase beta-subunit gene (*rpoB*) (Wahab et al. 2005).

2.1.7 Massively parallel pyrosequencing

High-throughput pyrosequencing, also called massively parallel pyrosequencing, (454 Life Sciences, a Roche company, Branford, CT, USA), has been developed and used mainly for large sequencing projects, as for sequencing of bacterial (Margulies *et al.* 2005) or human genomes (Wheeler *et al.* 2008). The genome of James D. Watson was sequenced in 7.4 fold redundancy in two months using newly released version of this technique (Wheeler *et al.* 2008). Massively parallel pyrosequencing has significantly higher throughput than the traditional Sanger-based sequencing method, generating 400 bp read length on average and 400 million bases data output in a ten-hour run with the current instrument (Genome Sequencer FLX system).

Massively parallel pyrosequencing is performed on a picotiter plate, a fibreoptic slide containing 3.6 million wells for pyrosequencing reactions. In massively parallel pyrosequencing, the sample DNA is sheared to fragments and ligated to universal oligonucleotide adaptors that contain sequences to which PCR and sequencing primers anneal and are thus used for amplification, sample preparation and sequencing. Two adaptors are ligated to each DNA fragment, and with help of the biotin in one of the adaptors, the fragments can bind to streptavidin coated bead under conditions (limiting dilutions) that favour one fragment per bead (Margulies *et al.* 2005).

The beads are added to water-in oil emulsion containing PCR reagents and the oil is divided to droplets, each containing one bead. The DNA fragments on the beads are each separately amplified in the oil droplets in emulsion PCR (emPCR). After the PCR, the PCR products on the beads are denatured and the sequencing primers are annealed to the single-stranded templates and each bead is loaded in a well of the picotiter plate. Thereafter, SSB and the large fragment of the *Bst* polymerase are bound to the DNA fragments on the beads. Then, smaller beads carrying luciferase and ATP sulfurylase are transferred into the wells of the picotiter plate. Pyrosequencing is performed by the sequential addition of substrates and nucleotides (dATP-α-S, dCTP, dGTP, dTTP) and a PPi standard. The unincorporated nucleotides are degraded by an apyrase wash between the nucleotide additions (Margulies *et al.* 2005).

Analysing more than one sample is possible in massively parallel pyrosequencing if specific barcodes are included in the universal adaptors. The use of eight-nucleotide barcode enables analysis of 1544 samples simultaneously (Hamady *et al.* 2008), but normally shorter tags suffice (Hoffmann *et al.* 2007).

2.1.8 Sanger sequencing

Chain termination method or the elegant Sanger sequencing has been the most widely used sequencing method for the last 30 years, also regarded as the gold standard. The principle of Sanger sequencing differs remarkably from pyrosequencing: Sanger sequencing is based on electrophoresis-based separation of PCR products having labelled dideoxynucleoside triphosphates (ddNTPs) as DNA chain terminators. In the original method, four different reactions were performed for a sample. Each reaction contained a single-stranded DNA template, a primer, a DNA polymerase, all the four dNTPs, and one radioactively labelled chain-terminating ddNTP. The labelled DNA fragments synthesized in the four reactions were separated by size using denaturing acrylamide gel electrophoresis and analysed by autoradiography. The nucleotide sequence was determined from the sizes of the DNA fragments deriving from the different reactions (Sanger *et al.* 1977).

In cycle sequencing, heat resistant DNA polymerase enables to use temperature cycling and linear amplification of the sequencing template using the sequencing primer (Murray 1989). The current widely used dideoxy sequencing known as fluorescent dye-terminator cycle sequencing is based on use of four fluorescently labelled ddNTPs. Each ddNTP is labelled with a different fluorescent dye. The fluorescent sequencing products are analysed by capillary electrophoresis, which separates the DNA fragments by size (smaller fragments move faster to the detector). At the detector, a laser beam excites the dye molecules of the DNA fragments of all analysed capillaries at the same time and the emission spectra is translated to a colour readout, called electropherogram.

2.2 General Microbiology

2.2.1 Microbial taxonomy

The taxonomy is a hierarchy of life established by Carl Linnaeus in 1735. In taxonomy, every living organism is included in a species, which in turn belongs to a genus, family, order, class, phylum or division, and domain. Prokaryotes are divided to two domains: the Archaea and Bacteria. The bacterial domain currently contains 29 accepted phyla (Euzéby 2011). In addition to the accepted phyla, 26 of 52 bacterial phyla described in 2003 were unaccepted; uncultured phyla that have been described based on 16S rRNA sequence data (Rappe and Giovannoni 2003). At present, taxa higher than genus are not commonly used even when describing new bacterial species. This is mainly because there are no clear rules for defining higher taxa (Garrity *et al.* 2007; Tindall *et al.* 2010).

Microbial taxonomists started to give microbial strains, pure cultures derived from a given specimen, genus and species designations in the 19th century. Only phenotypic characters were available at that time and the taxonomy was based on observable and measurable characteristics. As the methods developed, a bacterial species was defined as a bacterial group having a DNA-DNA hybridization (DDH) or reassociation value ≥70%. In addition, two strains belong to the same species if $\Delta T_{m,}$ difference of the melting temperature of a bacterial strain DNA and the melting temperature of the hybrid DNA of two strains, is ≤5 °C (Wayne 1988). In addition, the members of a species must possess at least one common diagnostic trait by which the members of this species can be distinguished from other species (Wayne 1988). DDH has several drawbacks, it is tedious, difficult to standardize and the results may depend on which strain is used as a probe and which as the target. In addition, the results of this method are not always extrapolatable; if isolates A and B as well as isolates B and C have DDH value of >70%, the value between A and C may well be <70%. This can naturally lead to controversial and complicated species designations and cumulative databases cannot be built on DDH data (Achtman and Wagner 2008; Richter and Rosselló-Móra 2009). Despite its drawbacks, DDH has remained the golden standard for genetically describing bacterial species.

16S rRNA sequencing was originally thought to substitute DDH, and it has been stated that isolates having ≤98.7% homologous 16S rRNA sequences always belong to different species because this degree of sequence difference in the 16S rRNA has been found to correlate with the DNA-DNA hybridization value of <70%. However, some species have >98.7% identical 16S rRNA sequences although their DNA-DNA reassociation values are <70% (Achtman and Wagner 2008). Therefore, isolates having less than 97% homologous 16S rRNA sequences may reliably be designated to different species but the isolates having >97% homologous 16S rRNA sequences cannot be reliably assigned to the same species based on the 16S rRNA sequence similarity only (Hanage *et al.* 2006). Although the degree of 16S rRNA sequence homology is useful in determining whether isolates belong to the same species or not,

new species cannot be described based on the 16S rRNA sequences alone due to the conservativeness of the 16S rRNA sequence (Richter and Rosselló-Móra 2009).

In 2002, the classification of new species was recommended to be based on MLSA, multilocus sequence analysis, the comparison of sequences of at least five housekeeping genes provided that there is congruence between the sequencing and DNA-DNA reassociation results (Stackebrandt *et al.* 2002). However, it is difficult to select which gene to sequence and the sequencing results may be confused by horizontal gene transfer (HGT) between species (Hanage *et al.* 2006; Milinovich *et al.* 2008).

Genome sequence comparison is ultimately the best method to describe species (Buckley and Roberts 2007), and the average nucleotide identity (ANI) between a given pair of genomes has potential to replace DDH. In ANI, genome sequences of two strains are aligned, and threshold of 95-96 % identical bases seems to correspond to the DDH of 70%. ANI can be performed also for draft genomes with preferably >50% coverage (Richter and Rosselló-Móra 2009). However, most of the genomes sequenced thus far are not type strains of the species, which poses a major obstacle in implementing the data for microbial taxonomy, and the type strain genomes should urgently be sequenced (Buckley and Roberts 2007; Richter and Rosselló-Móra 2009).

2.2.2 Bacterial ribosome

Ribosomes are the protein factories of all cells; they translate the sequence of mRNA molecules to the corresponding amino acid sequences. Ribosomes comprise two major parts, the large and small subunit, which in turn consist of protein and ribosomal RNA (rRNA) molecules. In the bacterial ribosome, the rRNA part of the small subunit contains approximately 1500 nucleotides and is called 16S rRNA based on its sedimentation velocity. Two rRNA molecules, 23S and 5S, are found in the large bacterial ribosomal subunit and contain approximately 3000 nucleotides in total. The small and large ribosomal subunits contain 20-21 and 31-35 proteins, respectively. The small ribosomal subunit binds to the mRNA and it contains the decoding center and controls the fidelity of translation. The large ribosomal subunit interacts with the tRNAs which bring their amino acids to the peptidyl transfer center (PTC) where the amino acids are bound to each other to form the peptide chain in the order which is coded in the mRNA. The protein exit tunnel, through which the peptide chain exits the ribosome, is also located in the large ribosomal subunit (Yonath 2005).

The structures of eukaryotic, bacterial, and archaeal ribosomes differ a little, and the subtle difference enables the antibiotics to hinder the action of bacterial but not eukaryotic ribosomes. Antibiotics interfere with the action of bacterial ribosome by binding to the decoding site, PTC, the protein exit tunnel, and other elements that are crucial for the function of the ribosome (Yonath 2005). The clinically most important antimicrobials binding to the ribosome are macrolides which will be discussed in chapter 2.4.1 of this thesis.

The sequences of the rRNA molecules, especially the sequence of the 16S rRNA, have been widely used for the identification of bacteria as the ribosomal sequences have

been found to be very useful in determining evolutionary relationships because the sequences act as molecular chronometers. Thus, the evolutionary position of all bacteria can be determined by sequencing its 16S rRNA. The ribosomal sequences are also useful because some regions of the molecules are highly conserved and others are more variable. (Woese 1987).

2.2.3 Human microbiome metagenomics

Only a small part of the human microbiota (or microbes in nature in general) is currently cultivable and is thus available for culture-based microbiological analysis. High-throughput sequencing has enabled the launch of a huge genome sequencing project, Human Microbiome Project, which aims to determine the human metagenome, combined sequences of the human host and the microbes belonging to the human microbiota by culture-independent methods (Turnbaugh *et al.* 2007).

Based on 16S rRNA sequence data, the human microbiota seems to be relatively conserved at the phylum level whereas there is host-specific difference in the genus, species and strain population levels and it was concluded that a common microbial core does not exist (Eckburg *et al.* 2005). However, in the study of Qin *et al*, the most common 57 species are present in ≥90% of individuals with genome coverage >1% and the abundance of these species varied extensively. The differences in the conclusions could be due to different study designs and methodology, and the earlier studies may have included only the most abundant species or sequences (Qin *et al.* 2010). Despite the varying views of the taxonomic composition of the microbiota, the current opinion is that the human microbiome seems to contain functionally similar genes that maintain the vital functions (Qin *et al.* 2010; Turnbaugh *et al.* 2009).

Composition and/or functionality of especially the human gut microbiota has been studied in various disorders, and the composition diseased gut has been found to differ from healthy (Chang *et al.* 2008; Frank *et al.* 2007). Interestingly, the gut microbiota has been found to be altered also in conditions that are not evidently related to gut as obesity (Turnbaugh *et al.* 2009) and autism (Parracho *et al.* 2005).

The science of metagenomics is currently at its pioneering stages, and for example, new bioinformatics tools have to be generated for managing the extensive amount of sequencing data before metagenomic analysis may become common. However, it is probable that metagenomic analysis of patient samples will become available sooner than we can anticipate. Once in routine use, metagenomic analysis will reveal all genetic components of the sample: the resistance determinants, virulence factors and current genotyping targets. Furthermore, well-annotated sequence data could give information about the clinical course of a patient with a given metagenome (Joseph and Read 2010).

2.3 The genus *Streptococcus*

Genus *Streptococcus* belongs to the phylum *Firmicutes*, class *Bacilli*, Order *Lactobacillales*, and family *Streptococcaceae* (Garrity *et al.* 2007). Streptococci are further divided to groups to which certain species are assigned. Microbial taxonomy is constantly redefined, and especially the grouping of streptococcal species has been constantly evolving.

The genus *Streptococcus* currently consists of more than 60 species and the number is increasing as new streptococcal, mainly alpha-haemolytic species are constantly described. Streptococci are both major pathogens and commensal organisms of humans and animals. They are gram-positive, catalase negative and facultatively anaerobic bacteria. Streptococcal cells are spherical or ovoid, less than 2 μm in diameter and form chains in liquid culture. Streptococci belong to the low G+C bacteria; their G+C content is between 34-46 mol% (Ruoff *et al.* 2003). Streptococci have traditionally been divided into beta- and alpha-haemolytic species. The beta-haemolytic isolates lyse erythrocytes completely; clearance of blood agar is distinctive. Alpha-haemolytic isolates produce a greenish darkening of the agar. Non-haemolytic streptococci are considered as alpha-haemolytic as incubation conditions may render a nonhaemolytic isolate alpha-haemolytic (Facklam 2002). The beta-haemolytic streptococci are generally easily identified based on the haemolysis and detection of Lancefield antigens (Lancefield 1933) whereas alpha-haemolytic isolates are more challenging to the clinical microbiology laboratory.

Streptococci are naturally competent for genetic transformation and therefore, HGT plays an important role in the evolution of these bacteria and affects also identification of streptococci (Delorme *et al.* 2007; Håvarstein *et al.* 1997; Kilian *et al.* 2008; Marri *et al.* 2006). HGT has been detected among the members of the *S. anginosus*, *S. mitis* and *S. salivarius* groups, and the homologous recombination happens within species of the same streptococcal group probably because sufficient sequence homology is required for recombination (Delorme *et al.* 2007; Hoshino *et al.* 2005; Jacobs *et al.* 2000; Kilian *et al.* 2008).

2.3.1 Beta-haemolytic streptococci

The most pathogenic of the beta-haemolytic streptococcal species is the group A streptococcus (GAS), *Streptococcus pyogenes*. It causes pharyngitis, impetigo and other, also invasive diseases. *S. pyogenes* is most easily identified by detection of the A antigen on the cell surface. However, some *S. anginosus* group members and *S. dysgalactiae* ssp. *equisimilis* have been found to infrequently carry the A antigen (Facklam 2002).

The group B streptococcus (GBS), *S. agalactiae* is the major cause of neonatal sepsis. In the 1970s it was noticed that transmission from colonized mother to the newborn is a major route of infection and that the risk of invasive disease of the newborn was proportional to the degree of maternal GBS colonization. For prevention of this disease, screening of pregnant women for carriage of this pathogen and subsequent

antimicrobial (penicillin) treatment have been established in the US. The disease burden has consequently significantly decreased (Baltimore 2007). Moreover, vaccination against GBS has been studied (Facklam 2002).

The C antigen is found in seven β-haemolytic species; *S. dysgalactiae* ssp. *equisimilis, S. equi* ssp. *equi, S. equi* ssp. *ruminatorum* (Fernandez *et al.* 2004), *S. equi* ssp. *zooepidemicus, S. anginosus, S. constellatus* ssp. *pharyngis,* and *S. phocae*. In addition, alpha-haemolytic *S. dysgalactiae* ssp. *dysgalactiae* also carries it. Consequently, the C antigen can only be used as an aid for identification of the isolates carrying it and the isolates should be further identified using other methods, as phenotypic tests. Human isolates with group C and G antigens forming large colonies belong to the species *S. dysgalactiae* ssp. *equisimilis* that may cause different kinds of severe infections as bacteraemia, endocarditis, meningitis, septic arthritis, and respiratory tract infections. The small colony forming isolates expressing A, C, F, or G antigens or being ungroupable belong to the former "*S. milleri*" group currently called *S. anginosus* group (Facklam 2002; Ruoff *et al.* 2003). *S. equi* ssp. *equi* causes strangles in horses and has not been isolated from human infections. *S. equi* ssp. *zooepidemicus* causes bovine mastitis, and most of the human infections are caused by consumption of contaminated dairy products. Human manifestations include septic infections and nephritis (Facklam 2002; Kuusi *et al.* 2006). *Streptococcus equi* ssp. *ruminatorum* has been detected in mastitis samples of small ruminants (Fernandez *et al.* 2004). *S. constellatus* ssp. *pharyngis* causes human pharyngitis. *S. phocae* was originally isolated from harbour seals and it has not been isolated from humans (Facklam 2002). *S. canis* carrying the Lancefield group G antigen has been isolated from animals, mostly dogs. *S. canis* has also been found to cause human sepsis, but the prevalence is unknown because the group G isolates are usually tested only for the Lancefield antigen and haemolysis (Facklam 2002).

S. porcinus is a beta-haemolytic streptococcus that carries the Lancefield group E, P, U, and V and four new antigens. *S. porcinus* has often been isolated from pyogenic infections of swine (Duarte *et al.* 2005) and also from female genitourinary tract but its clinical significance has not become clear as commercial GBS antibodies often cross-react with *S. porcinus* strains (Facklam 2002). *S. pseudoporcinus* was described in 2006, and the isolates were isolated from the genitourinary tract of women. The *S. pseudoporcinus* isolates were phenotypically similar to *S. porcinus* but could be differentiated as a separate species by 16S rDNA sequencing. Bekal *et al* have stated that human isolates seem to be *S. pseudoporcinus* whereas isolates from other sources are *S. porcinus* (Bekal *et al.* 2006). Consequently, the species found earlier in the genitourinary tract of women were probably also *S. pseudoporcinus* isolates. *S. pseudoporcinus* has also been found in a wound infection of a male patient (Mahlen and Clarridge 2009).

S. iniae, a beta-haemolytic streptococcus that does not possess any group antigen, was first found in freshwater dolphins and has later been found also in aquacultures of fish. It may also transfer to humans via the fish (Facklam 2002).

S. suis was formerly called Lancefield group R, S, and T streptococci, and these antigens are still useful for the identification of this bacterium. *S. suis* is a major porcine pathogen but it may also cause human infections by transferring to humans via a close contact with sick or carrier pigs. Mortality of meningitis, septicaemia, endocarditis, arthritis, and septic shock caused by *S. suis* is high in both pigs and humans (Lun *et al.* 2007). Some *S. suis* isolates are beta-haemolytic on agar plates containing horse blood, whereas all *S. suis* isolates are alpha-haemolytic when grown on sheep blood (Facklam 2002).

2.3.2 Alpha-haemolytic streptococci

The alpha-haemolytic streptococci are often divided to *S. pneumoniae* and viridans group streptococci, VGS (Ruoff *et al.* 2003). The VGS have been further divided into five groups based on the 16S rRNA sequence homology: *S. anginosus, S. bovis, S. mitis, S. mutans,* and *S. salivarius* group (Kawamura *et al.* 1995). In practice, however, isolates belonging to the *S. mutans* or *S. bovis* groups are not normally considered to belong to the VGS (Bishop *et al.* 2009).

The alpha-haemolytic streptococci are part of the human normal microbiota; they are present in the oral cavity, the gastrointestinal tract and the female genital tract. The commensal VGS have also been shown to inhibit growth of pathogenic bacteria. More specifically, *S. mitis, S. salivarius,* and *S. sanguinis* have shown inhibitory effect on *in vitro* colonization of periodontal pathogen *Aggregatibacter actinomycetemcomitans* (Norskov-Lauritsen and Kilian 2006; Teughels *et al.* 2007). It has also been found that the alpha-haemolytic streptococci of healthy children inhibit growth of otitis media pathogens *S. pneumoniae* and *Haemophilus influenzae* more efficiently than the alpha-haemolytic streptococci of children suffering from recurrent or secretory otitis media (Tano *et al.* 2000).

Since VGS colonize the mucous membranes, they are often considered as contaminants when isolated from patient samples, even from blood culture. However, VGS may indicate subacute bacterial endocarditis especially in patients with prosthetic heart valves. The incidence of viridans streptococcal infections is rising partly because the number of immuno-compromised neutropenic patients susceptible to this kind of infections is increasing (Hoshino *et al.* 2004; Kawamura *et al.* 1995; Ruoff *et al.* 2003). The most common causative agent of infective endocarditis (IE) is *Staphylococcus aureus* and the second most common are the oral streptococci which were considered the causative agent in 21% of 3784 IE cases (Moreillon and Que 2004). The most common viridans streptococcal species causing infections in neutropenic patients are *S. oralis, S. mitis*, and *S. salivarius* (Hoshino *et al.* 2004; Kawamura *et al.* 1995; Ruoff *et al.* 2003).

S. pneumoniae

S. pneumoniae is a major pathogen, it is the most common cause of community-acquired pneumonia and it causes also otitis media, sinusitis, meningitis, and endocarditis. However, oropharyngeal carriage of *S. pneumoniae* is common and the clinical significance of growth of *S. pneumoniae* especially from sputum is often difficult to estimate (Ruoff *et al.* 2003). *S. pneumoniae* is genetically (Kawamura *et al.* 1995) and phenotypically closely related to the members of the *S. mitis* group but it is usually not regarded as a member of *S. mitis* group. The major virulence factors of *S. pneumoniae* are its capsule, for which 90 serotypes are known, pneumolysin (Ply, a haemolytic cytolysin), cell wall autolysin (LytA), and many surface proteins that are thought to interact with host cells (Jedrzejas 2001).

Differentiation of *S. pneumoniae* from other, less pathogenic alpha-haemolytic streptococci is considered important. However, no golden standard exists for the identification of *S. pneumoniae* and various techniques are used for its identification. *S. pneumoniae* is most commonly differentiated from other alpha-haemolytic species by optochin susceptibility and bile solubility tests (Facklam 2002). Moreover, atypical optochin resistant and/or bile insoluble pneumococci are quite commonly detected especially in samples from nonsterile sites (Arbique *et al.* 2004; Kearns *et al.* 2000). It has been proposed that differentiation of *S. pneumoniae* from VGS should include at least colony morphology, haemolysis, and Gram stain morphology as well as either bile solubility, optochin susceptibility, or Phadebact coagglutination (Boule Diagnostics, Huddinge, Sweden) and information on the source of the isolate (Kellogg *et al.* 2001). Some isolates give atypical results in the aforementioned tests and their exact identification requires additional tests (Kellogg *et al.* 2001; Whatmore *et al.* 2000). The difficulty of differentiating *S. pneumoniae* from the members of the *S. mitis* group is demonstrated in the description of a new species *S. pseudopneumoniae*, the isolates of which were initially identified as *S. pneumoniae* (Arbique *et al.* 2004). Fortunately however, *S. pneumoniae* isolates isolated from blood are usually very homogenous and more easily identified than isolates from non-sterile sites (Clarridge 2004)

Accuprobe™ (Gen-Probe Inc, San Diego, CA, USA) is a widely used commercial hybridization test for the identification of *S. pneumoniae*. Accuprobe is based on the 16S rRNA sequence and it has been shown to have a sensitivity of 97.9% and specificity of 100% when compared to conventional identification methods (Lindholm and Sarkkinen 2004). However, *S. pseudopneumoniae* isolates have been found to be Accuprobe positive (Arbique *et al.* 2004).

Various PCR-based methods have been developed to identify *S. pneumoniae* preferably directly from patient samples. The PCR protocols are based on the detection of virulence genes of *S. pneumoniae* using conventional or real-time approach. The most often used targets are the pneumolysin gene *ply* and the autolysin gene *lytA* (Rudolph *et al.* 1993). However, *ply* has been detected also in *S. mitis* isolates possibly indicating a pathogenic role for this gene (Neeleman *et al.* 2004). In addition, both the *ply* and *lytA* genes have been detected in *S. pseudopneumoniae* (Arbique *et al.* 2004;

Neeleman *et al.* 2004). However, careful selection of PCR primer and probe sequences enabled the design of *S. pneumoniae*-specific *lytA* assay that did not amplify DNA from *S. pseudopneumoniae* or other tested *S. pneumoniae*-like streptococcal isolates (Carvalho *et al.* 2007).

S. mitis - *S. sanguinis* group

The species belonging to the *S. mitis* or *S. sanguinis* group are very closely related, and their identification is difficult both by phenotypic and genotypic methods. The *S. sanguinis* group can be distinguished from the *S. mitis* group by phenotypic tests (Facklam 2002) but not by analysis of the 16S rRNA sequence (Kawamura *et al.* 1995). Furthermore, the classification and nomenclature of the *S. mitis* group has caused considerable confusion. For example, the type strain of *S. mitis* has been changed and the former type strain has been designated to belong to *S. gordonii* (Kilian *et al.* 1989a). Furthermore, a term "biotype" for which no type strains are available, and the meaning of which seems to vary between studies has been widely used for especially *S. sanguinis* isolates (Kilian *et al.* 1989b). The *S. mitis* group comprises species *S. australis, S. cristatus, S. infantis, S. mitis, S. oralis, S. peroris*, and *S. pseudopneumoniae*. *S. sanguinis* group members are *S. sanguinis, S. parasanguinis,* and *S. gordonii* (Facklam 2002). In addition, *S. sinensis* that was originally isolated from blood of patients with infective endocarditis seems to belong either to the *S. anginosus* or to the *S. mitis* - *S. sanguinis* group by phenotypic and genetic bases (Woo *et al.* 2002).

Bacteria of the *S. mitis* group are often found as commensals in the oral cavity, the gastrointestinal tract and the female genital tract (Facklam 2002; Köhler 2007). The *S. mitis* group members can also transiently be found in the normal microbiota of skin and they can be contaminants when isolated from blood culture (Spellerberg and Brandt 2007). On the other hand, the streptococcal species most often causing endocarditis are *S. sanguinis, S. mitis, S. oralis,* and *S. gordonii,* which all are members of the combined *S. mitis* - *S. sanguinis* group (Facklam 2002; Ruoff *et al.* 2003). However, it is not possible to reliably connect any members of the *S. mitis* group to a certain infection yet (Facklam 2002).

S. mitis was described in 1906 (Andrewes and Horder 1906) which makes it the oldest species present in the group. *S. oralis* is an early-colonising organism of the oral cavity of infants and it is found also on adult teeth (Long and Swenson 1976). *S. cristatus* isolates have previously been called *S. sanguinis* biotype I, and this bacterial species has been isolated from dental plaques, periodontal abscesses and throats (Köhler 2007). *S. infantis* and *S. peroris* have been isolated from tooth surface and pharynx of infants. The first *S. peroris* strains were isolated from patients suffering from Kawasaki disease or acute pharyngitis (Kawamura *et al.* 1998). *S. orisratti* has been isolated from the mouths of rats that had been on high-sucrose diet (Zhu *et al.* 2000). The most recent additions to the *S. mitis* group are *S. australis* that was isolated from mouths of healthy children (Willcox *et al.* 2001), *S. oligofermentans* from dental plaques and saliva of Chinese caries-free patients with pharyngeal carcinoma (Tong *et al.* 2003), and *S. pseudopneumoniae* (Arbique *et al.* 2004).

S. sanguinis, S. parasanguinis, and *S. gordonii* belong to the *S. sanguinis* group (Facklam 2002). *S. sanguinis* causes severe bacteraemic infections as endocarditis but it is also present in the oral microbiota; colonization by it seems to begin at the same time as the first teeth erupt (Caufield *et al.* 2000). *S. parasanguinis* was originally isolated from human throat, blood and urine (Whiley *et al.* 1990). *S. sanguinis* and *S. parasanguinis* have been found to have a preventive effect on colonization of *S. mutans* and other microbes causing dental caries (Caufield *et al.* 2000; Tanzer *et al.* 2001). *S. massiliensis* isolated from human blood is the latest addition to the group. However, its designation was not straightforward as it was most closely related to the members of the *S. mutans* group by the 16S rRNA sequence whereas the *rpoB* and manganese-dependent superoxide dismutase gene (*sodA*) sequences indicated it to belong to the *S. sanguinis* group (Glazunova *et al.* 2006).

S. anginosus group

S. anginosus group members can be either beta- or alpha-haemolytic although majority of them are alpha-haemolytic and are grouped to VGS. This group currently includes three species and two subspecies: *S. anginosus, S. constellatus* ssp. *constellatus, S. constellatus* ssp. *pharyngis* and *S. intermedius*. Fortunately, the haemolysis seems to follow some rules in *S. anginosus* group streptococci: *S. intermedius* rarely seems to be beta-haemolytic, and on the other hand, *S. constellatus* tends to be most often beta-haemolytic of these species (Facklam 2002). *S. anginosus* group members are often isolated from oral abscesses and *S. anginosus* also from female genital infections. *S. intermedius* can be isolated from deep-seated abscesses, often from the liver or brain. *S. constellatus* is quite often isolated from thoracic sites and the respiratory tract although association with infection is less clear than with the other members of the *S. anginosus* group (Facklam 2002; Ruoff *et al.* 2003). *S. anginosus* group members have also been found to cause bacterial meningitis (Chang *et al.* 2002).

S. salivarius group

The *S. salivarius* group members *S. hyointestinalis*, *S. infantarius, S. salivarius, S. thermophilus*, and *S. vestibularis* are closely related to the *S. bovis* group members both phenotypically and based on their 16S rRNA sequences (Facklam 2002; Whiley and Beighton 1998). *S. hyointestinalis* isolates have been isolated from swine guts and were initially identified as *S. salivarius*, but the isolates were found to deserve a species status (Devriese *et al.* 1988; Facklam 2002). *S. infantarius* is currently listed in both *S. bovis* and *S. salivarius* groups because of its phenotypic characters: some of them are bile-esculin and bile negative, which excludes them from the *S. bovis* group. Nevertheless, they belong to the same species based on DNA - DNA reassociation results (Facklam 2002). *S. infantarius, S. salivarius,* and *S. vestibularis* have been isolated from humans and although *S. salivarius* is considered least pathogenic of the VGS, it is often isolated from various human infections (Corredoira *et al.* 2005). *S. vestibularis* can be isolated from the oral cavity and its pathogenicity to humans has not been established. *S. thermophilus* is the only streptococcus that is used in food production; it is commonly used with *Lactobacillus bulgaricus* as a starter culture in

dairy industry in producing fermented products as yogurt and Mozzarella cheese. These lactic acid bacteria also have important probiotic effects in relieving the symptoms of lactose intolerance, for example (Delorme 2008). *S. thermophilus* has been considered a subspecies of *S. salivarius* but the differences in their genetic and biochemical characteristics confirm their statuses as separate species (Schleifer *et al.* 1991).

S. bovis group

The *S. bovis* group has traditionally been divided to biotypes; biotype I strains are mannitol-positive and produce glucan from sucrose, whereas biotype II strains are mannitol-negative and do not produce glucan. The strains with biotype II.1 ferment starch, melibiose and glycogen. The strains of biotype II.2 ferment trehalose and produce β-glucuronidase and β-galactosidase. The biotypes are still commonly used for the identification of clinical strains (Ruoff *et al.* 2003).

The nomenclature and taxonomy of the *S. bovis* group has undergone major revisions. Species currently belonging to *S. bovis* or group D streptococci are *S. alactolyticus, S. bovis, S. equinus*, *S. gallolyticus* ssp. *gallolyticus S. gallolyticus* ssp. *pasteurianus*, *S. gallolyticus* ssp. *macedonicus, S. infantarius,* and *S. lutetiensis* (Köhler 2007; Schlegel *et al.* 2003). The *S. bovis* group has been found to contain six DNA homology groups by DDH analysis (Farrow *et al.* 1984). *S. bovis* and *S. equinus* isolates are mainly bovine isolates of biotype II.1 and belong to the DNA group 1 or rarely to the DNA group 3 that contains atypical *S. bovis* and *S. equinus* isolates. *S. bovis* and *S. equinus* are actually a single species, the name of which should be *S. equinus* as it has nomenclatural priority. However, *S. bovis* is still widely used in literature (Schlegel *et al.* 2003). In addition, no human *S. equinus* isolates have been found and consequently, all human isolates formerly called *S. bovis* I and II.2 have been suggested to be identified as *S. gallolyticus,* that forms the DNA group 2 (Facklam 2002; Schlegel *et al.* 2000). Bacteraemia caused by *S. gallolyticus* ssp. *gallolyticus* is associated with malignancies of the gastrointestinal tract (Herrero *et al.* 2002; Tjalsma *et al.* 2006) and patients with *S. gallolyticus* ssp. *gallolyticus* bacteraemia are usually examined for colon disease. *S. gallolyticus* ssp. *pasteurianus* are isolates formerly called *S. bovis* biotype II.2 and they have been found to cause meningitis and have been isolated mainly from human cerebrospinal fluid (Poyart *et al.* 2002). *S. gallolyticus* ssp. *macedonicus* was initially isolated from naturally fermented Greek Kasseri cheese (Schlegel *et al.* 2003; Tsakalidou *et al.* 1998) and has later been detected from other dairy products (Schlegel *et al.* 2003).

DNA group 4 includes *S. infantarius* that contains isolates previously identified as *S. bovis* biotype II.1. *S. infantarius* was initially divided to two subspecies; *S. infantarius* ssp. *infantarius* and *S. infantarius* ssp. *coli* (Schlegel *et al.* 2000). The type strain of *S. infantarius* ssp. *infantarius* was isolated from faeces of an infant and other isolates have been obtained from clinical specimens of humans or food products (Schlegel *et al.* 2000). *S. infantarius* ssp. *coli* have been isolated from humans: urine and blood of adults and faeces of infants (Schlegel *et al.* 2000). However, representatives of these subspecies were found to be so distantly related that they

cannot belong to the same species and *S. infantarius* ssp. *coli* isolates were suggested to be called *S. lutetiensis* (Poyart *et al.* 2002).

DNA group 5, *Streptococcus saccharolyticus* has been transferred to the genus *Enterococcus* and is currently called *Enterococcus saccharolyticus. S. alactolyticus*, DNA group 6, was described in 1984 and was first isolated from the intestines of pigs and faeces of chicken. *S. intestinalis*, also found in pig colons (Robinson *et al.* 1988), was found to be indistinguishable from *S. alactolyticus* and the isolates are collectively called *S. alactolyticus* (Vandamme *et al.* 1999).

S. mutans group

S. mutans, S. criceti, S. devriesei, S. downei, S. ferus, S. hyovaginalis, S. macacae, S. orisuis, S. ratti and *S. sobrinus* belong to the *S. mutans* group. *S. mutans* and *S. sobrinus* are the most common pathogens isolated from caries lesions or dental plaque and they are therefore also considered members of normal oral microbiota (Coykendall 1989; Ruoff *et al.* 2003). *S. mutans* has also been isolated from blood of patients with infective endocarditis (Gauduchon *et al.* 2003) and from diseased heart valve and atheromatous plaque lesions (Nakano *et al.* 2006). In one study, *S. mutans* was found both in oral and heart valve specimens of a patient, and the DNA fingerprints of the isolates were not identical. However, these results do not exclude the possibility of identical oral and heart valve *S. mutans* strains due to the small number of strains analysed in the study, 18 isolates from dental plaque and seven from heart valve (Nomura *et al.* 2006).

S. ratti and *S. cricetus* were originally isolated from laboratory animals but it is thought that the animals had received these bacteria from humans (Coykendall 1989). *S. macacae* is similar to *S. mutans* and it was originally isolated from the dental plaque of macacae monkeys (Beighton *et al.* 1984; Coykendall 1989). In addition, *S. downei* has been found in dental plaque of monkeys (Köhler 2007).

Many of the species of the *S. mutans* group have been found in mouths of various animals; *S. devriesei* has been isolated from teeth of horses and it may have cariogenic properties similar to *S. mutans* (Collins *et al.* 2004; Lundström *et al.* 2007). The most recent addition to the group, *S. orisuis,* was isolated from the oral cavity of a pig (Takada and Hirasawa 2007). *S. ferus* was originally isolated from the mouths of wild rats and this species was placed to the *S. mutans* group based on phenotypic criteria. However, *S. ferus* was found to be only distantly related to the other streptococci of the *S. mutans* group based on DNA-DNA reassociation studies and *sodA* and 16S rRNA sequences (Whatmore and Whiley 2002), and it seemed to be more closely related to the members of the *S. mitis* group (Kawamura *et al.* 1995). *S. hyovaginalis* was isolated from genital tracts of female swine and is probably part of their normal vaginal microbiota (Devriese *et al.* 1997). Also this species is included in the *S. mutans* group based on its phenotypic characteristics (Facklam 2002).

2.3.3 Ungrouped streptococci

Some of the described streptococcal species have not been deposited to a certain group (Table 1).

Table 1. Streptococcal species not designated to a certain group.

Species	Haem.[a]	Source	Reference
S. acidominimus	α	Vagina and udder of cows, milk, rarely human wounds or abscesses	(Ayers and Mudge 1922; Finkelstein *et al.* 2003)
S. caballi	α	Hindgut of horse with oligofructose-induced laminitis	(Milinovich *et al.* 2008)
S. castoreus	β	European beaver (*Castor fiber*)	(Lawson *et al.* 2005b)
S. dentirousetti	α	Tooth of a bat of the genus *Rousettus*	(Takada and Hirasawa 2008)
S. entericus	α	Cattle intestine	(Vela *et al.* 2002)
S. gallinaceus	α	Septicaemic adult broiler parents, human bacteraemia	(Balm *et al.* 2006; Collins *et al.* 2002)
S. halichoeri	-[b]	Seals	(Lawson *et al.* 2004)
S. henryi	α	Hindgut of horse with oligofructose-induced laminitis	(Milinovich *et al.* 2008)
S. ictaluri	-	Channel catfish *Ictalurus punctatus* broodstock	(Shewmaker *et al.* 2007)
S. intestinalis	α	Faeces of pigs, parakeet, dog and chicken, intestines of pigs	(Robinson *et al.* 1988; Vandamme *et al.* 1999)
S. marimammalium	-, β[c]	Seals	(Lawson *et al.* 2005a)
S. minor	α	Tonsils of dogs, a cat and a calf, anal swabs and faeces of dogs	(Vancanneyt *et al.* 2004)
S. ovis	α	Abscesses of sheep	(Collins *et al.* 2001)
S. parauberis	α?	Milk, mastitis sample	(Williams and Collins 1990)
S. pleomorphus[d]		Chickens, turkeys and ducks	(Barnes *et al.* 1977)
S. pluranimalium	α	Mastitis, genital and respiratory tracts of various animals	(Devriese *et al.* 1999)
S. thoraltensis	α	Vaginal fluid and intestine of swine	(Devriese *et al.* 1997)
S. uberis	α?		(Diernhofer 1932)
S. urinalis	-	Human cystitis	(Collins *et al.* 2000)

[a] Haemolysis
[b] Nonhaemolytic
[c] Nonhaemolytic on isolation, became β-haemolytic after 3 days
[d] Strictly anaerobic.

2.3.4 Identification of streptococci

Commercial identification systems mainly based on various enzyme tests are suitable for identification of common streptococci but often fail to identify alpha-haemolytic isolates. This is partly due to the new species that are not included in the databases of these tests (Hoshino *et al.* 2005). However, identification of also alpha-haemolytic isolates is important to better determine the different disease associations of these species and because different species seem to have different antimicrobial resistance profiles. However, it may well turn out that identification to the group level is sufficient for patient management, after all (Facklam 2002; Teng *et al.* 1998).

The first DNA-based identification methods were hybridization methods but even with them it became evident that VGS could not be reliably identified using phenotypic methods (Ezaki *et al.* 1988). The use of molecular methods, especially DNA-DNA reassociation has changed the taxonomy of streptococcal species, for example *S. bovis* has been included to the VGS (Ruoff *et al.* 2003), and has led to an increase in streptococcal species (Daley *et al.* 2005).

During the past decade, several DNA sequence-based methods for the identification of streptococci have been developed. The most often used target for bacterial identification has been the 16S rDNA, but its discrimination power has been argued to be too low for identification of closely related species, such as streptococci belonging to the *S. mitis* or *S. sanguinis* group. Consequently, methods based on more variable protein-coding genes have been developed: the targets include for example *rpoB*, the beta subunit of the RNA polymerase, (Drancourt *et al.* 2004), *sodA* (Poyart *et al.* 1998; Poyart *et al.* 2002), the *groESL* genes encoding 10-kDa (GroES) and 60-kDa (GroEL) heat shock proteins (Teng *et al.* 2002). In addition, 16S-23S rRNA gene intergenic spacer (Chen *et al.* 2004) and *rnpB*, the RNA subunit of endoribonuclease P have been used (Innings *et al.* 2005; Täpp *et al.* 2003). However, even these more variable targets fail to differentiate some streptococcal species. As an example *S. mitis* and *S. oralis* cannot be differentiated using the *groESL* gene due to intra-species variation in the sequence of the gene (Teng *et al.* 2002). When phylogenetic analysis is performed based on the sequences of the different targets, the streptococcal species usually group similarly to the 16S rRNA sequence (Teng *et al.* 2002). However, some species group differently with the different genes as mutation rates of the genes are different and HGT seems to occur more often in some genes than in others (Hoshino *et al.* 2005; Håvarstein *et al.* 1997). In addition, all these genes are not present in every organism although they are so-called house-keeping genes (Hoshino *et al.* 2005).

For determination of the most reliable target for the identification of streptococci, partial sequences of the *ddl, gdh, rpoB*, and *sodA* sequences were concatenated and compared to each other by Hoshino *et al* in 2005. All the analysed strains could be designated to a species using the concatenated sequences whereas analysis of the single sequences did not always result in the same tree topology as the concatenated sequence. The most deviating sequences were found among the *ddl* and *rpoB*

sequences whereas the *sodA* sequence generated similar results to the concatenated sequence.

In the relatively small study material of Hoshino et al, HGT seemed to occur most often within the *S. anginosus* group whereas it was rarer in *S. mitis* group although it was quite common in *S. sanguinis* (Hoshino *et al.* 2005). Due to HGT, use of multi-locus sequencing is recommended for the identification of streptococci but as it is not feasible, the best alternative in the opinion of Hoshino *et al.* would be partial analysis of the *sodA* gene (Hoshino *et al.* 2005) that has been found useful also in other studies (Garnier *et al.* 1997). However, *S. pneumoniae* isolates clustered with *S. mitis* isolates with partial (366 bp) *sodA* sequence, but all the *S. pneumoniae* isolates had specific base differences that could be used for the differentiation of *S. pneumoniae* from *S. mitis*. Interestingly, considerable intraspecies sequence variation, maximally 31 bases, was found within the region analysed. Nonetheless, the *S. pneumoniae* sequences were much more conserved as 11/16 sequences were identical and maximally two-base differences were found among all the studied *S. pneumoniae* strains (Kawamura *et al.* 1999).

Pyrosequencing of the *rnpB* gene has been applied for the identification of streptococci. The assay is based on analysis of 30 bp sequences of two variable regions of the gene. The pyrosequencing results of 113 isolates were compared to the results of commercial VITEK 2 and Rapid ID 32 Strep (RID32) (Biomérieux, Marcy L'Étoile, France) methods and the results of 58 isolates were identical with all three methods, 85 isolates had concordant results with pyrosequencing and VITEK 2 and 88 with pyrosequencing and RID32 (Innings *et al.* 2005).

MALDI-TOF MS analysis is an emerging technique for bacterial identification although it has been first proposed over 30 years ago (Anhalt and Fenselau 1975; Seng *et al.* 2009). This inexpensive, effective and rapid method has been used for the identification of bacteria that are hard to identify, including alpha-haemolytic streptococci. MALDI-TOF MS is based on the analysis of the protein mass spectra of bacterial pure cultures (Bizzini *et al.* 2011; Doern and Burnham 2010). The MS spectra generated from alpha-haemolytic streptococcal isolates belonging to same streptococcal group were more similar to each other than spectra of isolates that belonged to different streptococcal groups. Furthermore, the identification results generated by MALDI-TOF MS were generally in agreement with the results obtained by 16S rRNA sequencing (Friedrichs *et al.* 2007). MALDI-TOF seems to be a promising method for the identification of streptococcal isolates but it may not be the solution to the current problems regarding species-level identification of *S. pneumoniae* and *S. mitis* group isolates (Doern and Burnham 2010; Seng *et al.* 2009).

2.4 Antimicrobial resistance

Antimicrobial resistance is a major global health concern, and the situation is most likely getting worse due to increasing wealth and wider availability of antimicrobial drugs in developing countries. In addition, travelling and immigration spread the resistant isolates at an unforeseeable rate. Furthermore, there are no significant new antimicrobials that are expected to be released from the pipeline. Moreover, even if new antimicrobials were coming, bacteria would very likely acquire resistance against them soon after its launch in clinical use. Consequently, the quality and quantity of antibiotic use must be improved to slow the emergence and transmission of resistant isolates (Gould 2010). In addition, better diagnostic tools are needed to guide antimicrobial prescriptions in regards to increased incidence of resistant isolates that circulate in the society and the ever increasing variety of resistance elements found in these isolates.

Bacteria can prevent the action of most antimicrobial agents simply by keeping the antimicrobials outside of their cell walls. In gram-negative bacteria, the outer membrane protects the microbe from the antimicrobial agents. The larger, hydrophobic antimicrobial agents such as macrolides have to diffuse through the membrane and smaller, hydrophilic agents as β-lactams have to find a suitable pore, specific size-excluding channel, to go through the membrane. Gram-negative bacteria are intrinsically resistant to penicillin G because their outer membrane does not allow the drug to go to the periplasmic space (Delcour 2009; Martinez-Martinez 2008). This kind of intrinsic resistance is not clinically important as intrinsic resistance of clinically important pathogens is known before routine use of an antibiotic is launched. However, the properties of the bacterial cell membrane, especially loss of pore-forming porins, can affect the antimicrobial resistance spectrum of bacteria that have more specific resistance determinants (Delcour 2009; Martinez-Martinez 2008).

The first resistant clinical bacterial isolates were detected soon after the implementation of the first antimicrobial agent, penicillin. The rapid emergence of resistance was possible because the resistance determinants were present in the bacterial population already before implementation of the antimicrobial agent. When the use of penicillin was started, strong selection pressure was applied on the bacteria and the resistant isolates survived and resistance started to spread. Antimicrobial resistance is mainly spread by HGT and transmission of successful bacterial clones. In addition, bacteria acquire resistance mutations under selection pressure. The mechanism of antibiotic resistance can generally be divided in three: drug efflux, target site modification, and inactivation of the drug.

The standard method for determining antimicrobial resistance in bacterial isolates is to grow the bacteria in presence of a given antimicrobial drug and to determine whether growth of the bacteria is inhibited or not. Minimal inhibitory concentration (MIC) is determined by incubating bacteria on agar plates or in solutions containing certain concentrations of an antimicrobial, determining the lowest antibiotic concentration that is sufficient to inhibit growth of the bacteria. In routine laboratories, antimicrobial resistance is often determined by applying antibiotic disks on an agar plate not

containing antimicrobials and determining the inhibition zone around the antibiotic disc. Particularly in research laboratories, DNA-based methods are widely used to determine the mechanism or genetic elements causing the resistance phenotypes and to study the epidemiology, prevalence, and spread of these elements.

2.4.1 Macrolide antibiotics

Macrolides compose of a 12-16 member macrolactone ring to which various amino sugars are attached. Erythromycin (Figure 4) is a natural secondary metabolite of *Saccharopolyspora erythraea* whereas other macrolides are semisynthetic. The clinically useful macrolides consist of a 14-16 - member lactone ring that is substituted with at least two neutral or amino sugars (Vester and Douthwaite 2001). Macrolides are normally active against streptococci, staphylococci and enterococci as well as against mycoplasma, some rickettsia, chlamydia (Alvarez-Elcoro and Enzler 1999) and mycobacteria except for the *Mycobacterium tuberculosis* complex (Buriankova *et al.* 2004).

Macrolides block bacterial protein synthesis by binding to the upper part of a pocket in the protein exit tunnel found in the vicinity of the PTC in the large subunit of the bacterial ribosome. By binding to the pocket, erythromycin and other 14-member-ring macrolides block the growth of the nascent polypeptide chain thereby causing premature arrest of elongation and dissociation of the peptidyl-tRNA from the ribosome (Mankin 2008; Schlünzen *et al.* 2001; Tenson *et al.* 2003; Yonath 2005). Macrolides also prevent new ribosomal complexes from assembling, which leads to depletion of ribosomes in the bacterial cell (Vester and Douthwaite 2001; Yonath 2005). The lactone ring of macrolides binds to the protein exit tunnel wall primary by hydrophobic interactions with the residues 2058 and 2059 (*E. coli* numbering*)* of the bacterial 23S rRNA (Mankin 2008). In addition to macrolides, the nascent protein tunnel pocket is the target of ketolide and streptogramin B antibiotics. However, the specific interactions of different antimicrobials binding to the tunnel likely vary (Mankin 2008; Yonath 2005).

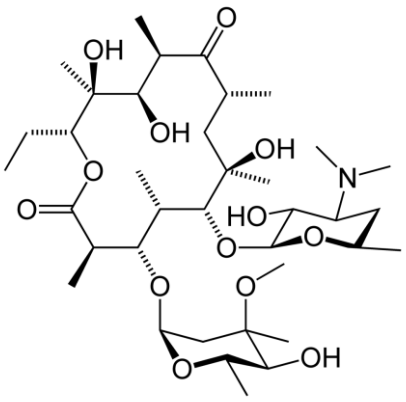


Figure 4. Structure of erythromycin.

Methylases and efflux

Isolates resistant to erythromycin started to emerge shortly after introduction of erythromycin in clinical use in 1952. Soon it was noticed that the isolates were resistant to all macrolides and to unrelated lincosamide and streptogramin B drugs; they had the MLS_B resistance phenotype. This resistance phenotype was first described in *Staphylococcus aureus*, and it is caused by an Erm (erythromycin ribosome methylation) methylase that specifically dimethylates the N-6 position of the adenosine residue at the position 2058 of the bacterial 23S rRNA. A number of different Erm methylases have been described in different bacteria (Roberts *et al.* 1999) and expression of the *erm* gene may be constitutive or induced by a macrolide drug (Weisblum 1995). Due to the dimethylation of the residue 2058 of the 23S rRNA, macrolides are not able to bind to the protein exit tunnel due to steric hindrance and missing hydrophobic interactions between the lactone ring and the bacterial ribosome (Schlünzen *et al.* 2001).

Low-level resistance to 14- and 15-member-ring macrolides, the M phenotype is caused by macrolide efflux (Mef) pumps. The first *mef* gene, *mef*(A) was first described in *S. pyogenes.* This membrane-bound efflux pump uses proton-motive energy to keep the macrolide concentration low in the bacterial cell (Clancy *et al.* 1996). The *mef* gene first found in *S. pneumoniae* was initially called *mef*(E). Because *mef*(A) and *mef*(E) elements share 90% of their nucleotide sequences, these elements were proposed to be collectively called *mef*(A) (Roberts *et al.* 1999). It has later been found that these resistance genes reside in different genetic elements (Santagati *et al.* 2003; Santagati *et al.* 2000) and thus, the differentiation of these elements is again warranted (Del Grosso *et al.* 2002). Currently *mef*(A) is common both in *S. pyogenes* and *S. pneumoniae* whereas *mef*(E) is common in all streptococci except *S. pyogenes*. Also other uncommon *mef* genes exist in streptococci (Varaldo *et al.* 2009). The *erm* and *mef* macrolide resistance genes can be detected by PCR (Clancy *et al.* 1996; Sutcliffe *et al.* 1996).

Ribosomal mutations

The macrolide resistance mutations disrupt the hydrophobic interactions between the ribosomal protein exit tunnel and macrolides. The most important resistance mutations occur at the positions 2058 and 2059 (*E. coli* numbering) of the 23S rRNA that are located in the region V of the 23S rRNA and the protein exit tunnel of the ribosome (Tait-Kamradt *et al.* 2000b). Macrolide resistance mutations can be induced to the rRNA genes *in vitro* by culturing bacteria in media containing low concentrations of macrolides (Canu *et al.* 2002) and resistance mutations often develop during macrolide therapy (Tait-Kamradt *et al.* 2000b; Vester and Douthwaite 2001).

Macrolide resistance mutations in the 23S rRNA are more common in species harbouring only one or two ribosomal genes, such as mycobacteria or *Helicobacter pylori* whereas species having more ribosomal genes as enterococci, streptococci or staphylococci more often become resistant by acquiring the resistance genes *erm* or *mef* (Vester and Douthwaite 2001). However, mutations conferring macrolide

resistance have been detected at the positions 2058-2059 of the 23S rRNA in also bacterial species harbouring more than two ribosomal alleles, for example in *S. pneumoniae, S. pyogenes*, *Bordetella pertussis,* and campylobacteria (Bartkus *et al.* 2003; Engberg *et al.* 2001; Jalava *et al.* 2004; Tait-Kamradt *et al.* 2000a; Tait-Kamradt *et al.* 2000b). The number of mutated alleles has been shown to correlate with the level of resistance in bacteria harbouring more than one 23S sequence (Tait-Kamradt *et al.* 2000a).

The exact position of the mutation in the protein exit tunnel has effect on the resistance: if the mutation of a *S. pneumoniae* strain is at position 2058, the mutant confers similar resistance pattern to Erm methylation, strong resistance to macrolide, lincosamide, and streptogramin B antibiotics (Tait-Kamradt *et al.* 2000b). The similar resistance phenotypes of an A2058G mutation and Erm methylation are expected because both mechanisms increase the size of the ribosomal nucleotide A2058 so that macrolides are not able to bind to the protein exit tunnel (Vester and Douthwaite 2001; Yonath 2005). Mutation at the position 2059 of the 23S rRNA causes ML (macrolide-lincosamide) resistance phenotype in pneumococci (Tait-Kamradt *et al.* 2000b).

The mutations in the 23S rRNA can be detected using real-time PCR (Lascols *et al.* 2003; Schabereiter-Gurtner *et al.* 2004; Wolff *et al.* 2008), restriction fragment polymorphism (Vacher *et al.* 2003) or denaturing high performance chromatography (Canu *et al.* 2004). Quantitative detection of the mutations by conventional sequencing is possible but requires allele-specific amplification (Jalava *et al.* 2004; Tait-Kamradt *et al.* 2000b).

Mutations in the L4 and L22 ribosomal proteins have also been shown to cause macrolide resistance in several bacterial species including *S. pneumoniae* (Pihlajamäki *et al.* 2002b; Tait-Kamradt *et al.* 2000a; Tait-Kamradt *et al.* 2000b).

2.4.2 Beta-lactam antibiotics

The first beta-lactam, penicillin was discovered in 1929 by Alexander Fleming. However, due to difficulties in production and purification of this substance, the first antimicrobial agent, penicillin G was introduced into clinical use in 1941. Since then, several beta-lactam drugs have been developed, and they are the most often used antimicrobial agents worldwide. The beta-lactam antibiotics by definition contain the beta-lactam ring, a four-atom ring structure consisting of three carbon atoms and one nitrogen atom (Figure 5). The beta-lactams act by inhibiting the synthesis of bacterial cell wall by inactivating the so-called penicillin binding proteins, PBPs that are essential in bacterial cell wall peptidoglycan synthesis (Wright 1999). Due to the bacteria-specific target, the toxicity of beta-lactams to humans is low.

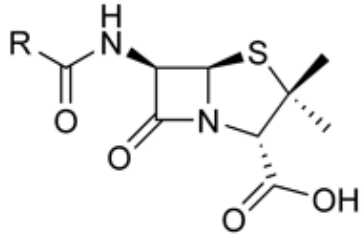


Figure 5. Structure of the penicillin core. The β-lactam ring consisting of three carbon and one nitrogen atom is found in the middle of the β-lactam molecules.

Beta-lactams are active against wide variety of bacteria; and they are used in primary care for treatment of several types of infections as otitis media, sinusitis, tonsillitis and skin infections (Rautakorpi *et al.* 1999). Beta-lactams are used for treatment of many kinds of infections, urinary tract infection, respiratory infections as well as severe, life-threatening infections as sepsis. Due to the development of resistance to beta-lactams, penicillins have been replaced by other antimicrobials in some indications (Wright 1999).

Beta-lactamases

Resistance to beta-lactams is mainly mediated by beta-lactamases, enzymes that inactivate beta-lactams by breaking the beta-lactam ring. The first beta-lactamase was found in 1940, before penicillin was in clinical use (Abraham and Chain 1940), and some bacteria as *Enterobacteriaceae* are naturally resistant to beta-lactams since they harbour a chromosomal beta-lactamase gene (Bradford 2001). Numerous types of beta-lactamases with different enzymatic and molecular properties have been found in different bacteria. The rapid spread of β-lactamase genes is due to the occurrence of these genes in plasmids; the first plasmid-mediated gene, TEM-1, was found in the early 1960s. The name of the gene derives from the name of the patient, Temoniera, from whom the first TEM-1 producing *E. coli* strain was isolated. Within a few years, the plasmid-mediated TEM-1 β-lactamase spread worldwide and is at present commonly found in penicillin-resistant *Enterobacteriaceae, Haemophilus influenzae,* and *Neisseria gonorrheae* (Bradford 2001).

SHV (for sulfhydryl variable) is another important β-lactamase found mainly in *K. pneumoniae* and *E. coli*. SHV-1 and TEM-1 have 68% identical amino acid sequences and their structures are similar (Jacoby and Munoz-Price 2005). The SHV-1 and TEM-1 enzymes hydrolyze ampicillin, amoxicillin, and other penicillins as well as early generation cephalosporins (Paterson 2006). The first bacterial isolate expressing an ESBL (extended spectrum beta-lactamase), SHV-2, was found in the 1980s (Kliebe *et al.* 1985). SHV and TEM ESBLs derive by amino acid substitutions from the SHV-1 and TEM-1 enzymes and are able to hydrolyse also third generation cephalosporins and aztreonam (Paterson 2006). The amino acid substitutions leading to an ESBL phenotype occur at a limited number of positions in both enzymes, and the substitution (s) leading to the ESBL phenotype are usually located at the active site of the enzyme or at its vicinity. On the other hand, these substitutions expose the active site of enzymes to the activity of β-lactamase inhibitors such as clavulanic acid. Consequently, resistance to both beta-lactams and β-lactamase inhibitors is fortunately rare (Jacoby and Munoz-Price 2005).

At present, hundreds of different beta-lactamases have been described and they are categorized to four major groups A-D based on their primary structure. Classes A, C, and D are homologous enzymes and they contain a serine residue at their active site. SHV and TEM belong to the class A enzymes for which Ambler *et al* (1991) have designated a standard numbering scheme. Class C enzymes, AmpC enzymes hydrolyze broad and extended spectrum cephalosporins but are not inhibited by beta-lactamase inhibitors, as clavulanic acid. The class D, OXA-type, enzymes confer resistance to

ampicillin and cephalothin, have high hydrolytic activity against oxacillin and cloxacillin and are poorly inhibited by clavulanic acid. Amino acid substitutions in OXA enzymes can also lead to the ESBL phenotype. Originally, OXA-type enzymes were found in *P. aeruginosa* but more recently, OXA type enzymes have been found in *Acinetobacter baumannii* isolates (Jacoby and Munoz-Price 2005; Pfeifer *et al.*).

Class B enzymes are metallo-β-lactamases and usually contain a zinc ion at their active site. Class B enzymes are capable of hydrolyzing carbapenem antibiotics that are resistant to most beta-lactamases and often considered as the best treatment choice when ESBL producers are present. In addition, class B enzymes are not inhibited by beta-lactamase inhibitors, as clavulanic acid (Jacoby and Munoz-Price 2005).

The SHV and TEM type beta-lactamases used to be the most common beta-lactamases worldwide. Currently however, the most common beta-lactamases in many European countries are of CTX-type (Coque *et al.* 2008). CTX is also a group A enzyme that usually has greater activity against cefotaxime than ceftazidime (Jacoby and Munoz-Price 2005).

The ESBLs are currently a significant clinical problem worldwide, and detection of ESBLs is complicated due to the heterogeneity of the enzymes, their variable activity against different drugs, their coexistence in clinical isolates, and variable expression of these enzymes. Consequently, detection of ESBLs in clinical laboratories may be difficult, patients may be at risk of receiving inappropriate treatment, and the ESBLs may be underreported which may lead to spread of resistant isolates. Furthermore, estimation of the clinical significance of an isolate carrying a specific ESBL gene is difficult since the specificity of hydrolysis, the level of expression and the presence of additional resistance factors cannot be determined by determining only the ESBL type of an isolate (Bush 2001).

SHV

The SHV gene of clinical isolates usually resides in plasmids. However, most *Klebsiella pneumoniae* isolates also carry a chromosomal SHV or a related beta-lactamase gene (Babini and Livermore 2000; Bradford 2001; Chaves *et al.* 2001) that is thought to be the ancestor of the plasmid-encoded SHV variants (Hæggman *et al.* 1997). The chromosomal SHV gene is usually not of ESBL type; the most common chromosomal variants are the SHV-1 (wild type) and SHV-11 that differs from the SHV-1 by a leucine to glutamine mutation at the position 35 (Lee *et al.* 2006).

The chromosomal SHVs may interfere with determination of the SHV type of an isolate. The PCRs for detection of SHV genes usually amplify both the chromosomal and the plasmid copy. Determining the exact SHV type from a duplex PCR product by traditional sequencing requires either very careful analysis of the electropherograms (al Naiemi *et al.* 2006) or cloning and sequencing the PCR product.

The mutations causing ESBL phenotype in SHV most often occur at the amino acid positions 238 and/or 240 of the so called Ambler numbering scheme (Ambler *et al.* 1991). In addition, mutations at positions 146, 156, 169, 179 and 205 have been shown to be associated with ESBL activity. Mutations at positions 238 and 179 seem to be critical for ESBL activity, and at least one of these mutations is found in the majority of ESBL type SHV enzymes. These amino acids reside at or at the vicinity of the active site of the enzyme (Gniadkowski 2008). In addition, mutation at position 35 is common among SHV variants but it does not alone cause an ESBL phenotype. Currently (February, 2010) 132 different SHV variants have been described and the sequence of 114 SHV types has been released. Of these, 75 contain a mutation at the position 35, 238 or 240 (44, 35 and 32 types have mutation at the positions 35, 238 and 240, respectively) (Jacoby and Bush 2010).

3 AIMS OF THE STUDY

The aim of this study was to develop molecular methods for the identification of bacteria and detection of resistance mechanisms

The specific aims of the study were:

1. To develop a pyrosequencing assay for the identification of (alpha-haemolytic) streptococci (I-II)

2. To develop a pyrosequencing method for the detection and quantification of major macrolide resistance mutations in the 23S rRNA (III)

3. To develop a pyrosequencing assay for detection of the most important mutations conferring ESBL phenotype in SHV enzyme (IV)

4 MATERIALS AND METHODS

4.1 Bacterial isolates (I-IV))

Streptococcal type and other strains deposited in culture collections used in this study are presented in Table 2. As the streptococcal species nomenclature constantly evolves, there are more streptococcal type strains available than there are currently accepted species. Of the species listed in Table 2, *S. infantarius* ssp. *coli* and *S. lutetiensis* are considered members of the same species, *S. lutetiensis* (Poyart *et al.* 2002) and *S. caprinus* is considered a member of *S. gallolyticus* ssp. *gallolyticus* (Sly *et al.* 1997), whereas *S. waius* is considered to belong to *S. gallolyticus* ssp. *macedonicus* (Manachini *et al.* 2002). In addition, *S. difficilis* is currently considered to belong to *S. agalactiae* (Kawamura *et al.* 2005), and *S. bovis* strains to belong to *S. equinus* (Schlegel *et al.* 2003). Published sequences of newly described species were added to the sequence database also after submitting study II. The newly described species whose sequences were added to the database were not closely related to the isolates of this study, and the added sequences did not affect the results of study II.

Table 2. Streptococcal strains whose combined v1+v2 16S rRNA sequences were included in the local sequence database of this study.

Type strain	Study[a]	Type strain	Study[a]
S. acidominimus LMG 17755^T	II	*S. ictaluri* ATCC BAA-1300^T	–[b]
S. agalactiae LMG 14694^T	II	*S. infantarius* ssp. *coli* CCUG 47831^T	II
S. alactolyticus CCUG 27297^T	II	*S. infantis* LMG 18720^T	II
S. anginosus DSM 20563^T	I, II	*S. iniae* CCUG 27303^T	II
S. australis CCUG 45919^T	II	*S. intermedius* DSM 20573^T	I, II
S. bovis LMG 8518^T	II	*S. lutetiensis* CCUG 46149^T	II
S. caballi 151^T	–[b]	*S. macacae* LMG 15097^T	II
S. canis LMG 15890^T	II	*S. marimammalium* M54/01/1^T	–[b]
S. caprinus LMG 15572^T	II	*S. massiliensis* 4401825^T	–[b]
S. castoreus M605815/03/2	–[c]	*S. minor* LMG 21735	–[b]
S. constellatus ssp. *constellatus* DSM 20575^T	I, II	*S. mitis* DSM 12643^T	I, II
		S. mutans CCUG 17824^T	II
S. constellatus ssp. *pharyngis* CCUG 46377^T	II	*S. oligofermentans* 2-4^T	–[c]
		S. oralis DSM 20627^T	I, II
S. criceti LMG 14508^T	II	*S. orisratti* CCUG 43577^T	II
S. cristatus DSM 8249^T	I, II	*S. orisuis* AB182324.1	–[b]
S. didelphis CCUG 45419^T	II	*S. ovis* LMG 19174^T	II
S. devriesei CCUG 47156^T	–[b]	*S. parasanguinis* DSM 6778^T	I, II
S. difficilis ATCC 51487^T	–[b]	*S. parauberis* LMG 12174^T	II
S. downei CCUG 24890^T	II	*S. peroris* LMG 18719^T	II
S. dysgalactiae ssp. *dysgalactiae* LMG 15885^T	II	*S. phocae* LMG 16735^T	II
		S. pleomorphus CCUG 11733^T	II
S. dysgalactiae ssp. *equisimilis* CCUG 36637^T	II	*S. pluranimalium* LMG 14177^T	II
		S. pneumoniae ATCC 49619	I, II
S. entericus CCUG 44616^T	II	*S. pneumoniae* DSM 11867	II

Continued on following page

Table 2 – continued

Type strain	Study[a]	Type strain	Study[a]
S. equi ssp. *ruminatorum* CECT5772^T	–[c]	*S. pneumoniae* LMG 14545^T	II
		S. porcinus LMG 15980 ^{T}	II
S. equi ssp. *equi* LMG 15886^T	II	*S. pseudopneumoniae* ATCC	–[c]
S. equi ssp. *zooepidemicus* LMG 16030^T	II	*S. pseudoporcinus* LQ 940-04^T	–[b]
		S. pyogenes ATCC 700294	II
S. equinus LMG 14897^T	II	*S. pyogenes* CCUG 4207^T	II
S. ferus LMG 16520^T	II	*S. ratti* CCUG 27642^T	II
S. gallinaceus CCUG 42692^T	–[c]	*S. salivarius* DSM 20560^T	I, II
S. gallolyticus ssp. *gallolyticus* CCUG 35224^T	II	*S. sanguinis* DSM 20567^T	I, II
		S. sinensis CCUG #0059#^{T}	II
S. gallolyticus ssp. *macedonicus* LMG 18488^T	II	*S. sobrinus* CCUG 25735^T	II
		S. suis CCUG 7984^T	II
S. gallolyticus ssp. *pasteurianus* CCUG 46150^T	II	*S. thermophilus* CCUG 21957^{T,d}	II[d]
		S. thoraltensis LMG 13593^T	II
S. gordonii DSM 6777^T	I, II	*S. uberis* LMG 9465^T	II
S. halichoeri CCUG48324^T	–[b]	*S. urinalis* LMG 19649^T	II
S. henryi 126^T	–[b]	*S. vestibularis* DSM 5636^T	I, II
S. hyointestinalis LMG 14579	II	*S. waius* 3/1	–[b]
S. hyovaginalis LMG 14710^T	II		

^{T} Type strain

[a] Strains analysed by pyrosequencing in study I and/or II are indicated by the roman numbers.

[b] Sequence of this type strain was retrieved from the GenBank and included in the local sequence database while writing this thesis.

[c] Sequence of this type strain was retrieved from the GenBank and included in the local sequence database of study II.

[d] *S. thermophilus* was not considered a realistic identification result in this study as its optimal growth temperature is 45 ºC and human isolates were studied in this study.

In study I, 172 isolates from the normal pharyngeal (n=151), nasopharyngeal (n=7) and conjunctival (n=3) microbiota of 28 elderly persons were analysed. The isolates were isolated from the patients before cataract surgery and the patients had not received any antimicrobial treatment during at least three months preceding the sampling.

In study II, thirty-nine of the pharyngeal isolates from five persons of study I, 99 alpha-haemolytic streptococcal blood culture isolates collected from 96 patients in the Turku University Hospital between 1993 and 2004, and 17 invasive *S. pneumoniae* strains isolated in Finland in 2002 were included.

In study III, 13 well characterized bacterial strains (Table 3) were used.

Table 3. Strains analysed in study III.

Strain	ERY MIC[a] (μg/ml)	Mutant 23S rRNA alleles[b]	Reference
S. pneumoniae ATCC 49619	0,125	0/4	(Pihlajamäki *et al.* 2002b)
S. pyogenes ATCC 700294	0,063	0/6	(Jalava *et al.* 2004)
S. pyogenes NI 4277	>256	5/6	(Jalava *et al.* 2004)
S. pneumoniae r581	128		
S. pneumoniae r506	256	2/4	(Pihlajamäki *et al.* 2002b)
S. pneumoniae r771	256	2/4	(Pihlajamäki *et al.* 2002b)
S. pneumoniae r1317	>512	3/4	(Pihlajamäki *et al.* 2002b)
S. pneumoniae r733	>512	4/4	(Pihlajamäki *et al.* 2002b)
S. pneumoniae 01-41	256		Littauer, 2005 #538}
C. jejuni 62	>256		
H. influenzae 286	>64		
M. avium H0812/96	>256[c]		
M. avium H0851/96	≤8[c]		

[a] ERY = erythromycin
[b] Previously known resistance mutations at the position 2058 or 2059 of the 23S rRNA alleles
[c] Azithromycin MIC

In study IV, SHV-positive *E. coli* (n=10) and *K. pneumoniae* (n=106) isolates collected in various Finnish hospitals were included. The isolates were originally sent to the National Public Health Institute for confirmation of the ESBL phenotype. The SHV type of 40 of the *K. pneumoniae* isolates could not be determined by cycle sequencing due to heterogeneous sequences at the Ambler (Ambler *et al.* 1991) positions 35, 238 and/or 240 (Nyberg *et al.* 2007).

4.2 Ethical issues

The collection of isolates from patients in study I was authorized by the Ethical Committees of the Hospital District of Southwest Finland and University of Turku. Clinical strains of the studies II-IV belong to the microbe collection of the National Institute for Health and Welfare.

4.3 Template preparation (I-IV)

The DNA used in pyrosequencing (I-IV) was extracted by a boiling method: bacterial colonies from agar plates were suspended in water and heat-inactivated at 95 ˚C for 10 min. From the mycobacterial isolates (III), DNA was liberated by heat-inactivation at 95 ºC for 45 min and subsequent sonication (RT, 15 min).

4.4 PCR (I-IV)

The primers used in PCRs, pyrosequencing and cycle sequencing reactions are listed in Table 4. All primers were HPLC purified and synthesized by Thermo Fisher Scientific (Waltham, MA, USA).

The PCRs were performed using various thermocyclers and cycling conditions, the details of which are available in the original publications. After the PCRs, the presence of the PCR products was verified on a stained agarose gel. The PCR products used in pyrosequencing should preferably have a strong visual band on the gel and they should not contain primer-dimers or nonspecific amplification products as they also bind to the sepharose beads and may give rise to nonspecific signals particularly if the PCR primer is used as the sequencing primer.

4.5 Pyrosequencing (I-IV)

In pyrosequencing the biotin-labelled PCR products attach to streptavidin coated sepharose beads (GE Healthcare, Buckinghamshire, UK) by agitation on a PCR plate using binding buffer (10 mM Tris-HCl, 2 M NaCl, 1 mM EDTA, 0.1% (v/v) Tween 20) for 5 min or longer. In the applications of this study, 20 µl PCR reaction and 3 µl streptavidin beads were used for each pyrosequencing reaction. After binding the PCR products to the beads, the PCR products were washed and rendered single-stranded with the Vacuum Workstation (Qiagen). Using the Vacuum Workstation, the beads were collected from the solution and attached to filters through which different solutions were aspirated. First, the beads and the PCR products were washed with 70% ethanol to remove Tween residues of the binding buffer. Thereafter, the PCR products were denatured using 0.2 M NaOH and then the beads were washed with 10 mM Tris-acetate and deposited into the wells of the sequencing plate containing primer solution. The primer solution (40µl) contained annealing Buffer, (20 mM Tris-acetate, 2 mM magnesium acetate, pH 7.6) and 15 pmol sequencing primer(s). Thereafter, the PCR products were heated at 80 °C for 2 min and cooled to room temperature in order to avoid formation of secondary structures, and to anneal the sequencing primers.

The run was programmed to the PSQ 96MA pyrosequencing instrument (Qiagen) which calculated the amounts of nucleotides and enzymes needed for the run based on the number of reactions and the dispensation orders. The required amounts of nucleotides were pipetted to the reagent cartridge and 5 µl of enzymes and substrates were usually pipetted manually to the wells to save reagents. The most often used dispensation orders of this study are listed in table 5. After starting a pyrosequencing run, the instrument first dispenses the enzymes and substrates to the sequencing wells and thereafter the dNTPs in the chosen dispensation order.

The pyrosequencing run may be monitored in real time. The software interprets the signals and presents the data as a sequence signal peaks or pyrogram. The analysis software performs base calling but when ambiguities occur, the sequences may be monitored visually.

Table 4. Primers used in this study.

Name	Sequence	Target	Position	Use	Reference
Bio-StrepV1For	Bio-AGTTTGATCCTGGCTCAGGACG	16S rRNA	-1- 22[a]	PCR_P[b]	I, II
StrepV1rev	TTACCTACGCGTTACTCACCCG	16S rRNA	117-96[a]	PCR_P[b] ,pyroseq[c]	I
StrepV2Rev	CAACTAGCTAATACAACGCAGGTC	16S rRNA	248 – 225[a]	PCR_P[b]	II
StrepV1RevV2	CCTACGCGTTACTCACCCGTTC	16S rRNA	114 – 93[a]	Pyroseq[c]	II
StrepV2RevV2	ACTAGCTAATACAACGCAGGTCCA	16S rRNA	246 – 223[a]	Pyroseq[c]	II
fd1mod2	AGAGTTTGATCMTGGCTCAG	16S rRNA	-3 – 20[a]	PCR_C[d]	II
rp2	ACGGCTACCTTGTTACGACTT	16S rRNA	1506 – 1486[a]	PCR_C[d]	II, Jalava *et al.* 1995
357f	TACGGGAGGCAGCAG	16S rRNA	337 – 351[a]	C_Seq[e]	II, Jalava *et al.* 1995
357r	CTGCTGCCTCCCGTA	16S rRNA	351 – 337[a]	C_Seq[e]	II, Jalava *et al.* 1995
533f	GTGCCAGCAGCCGCGGTAA	16S rRNA	510 – 528[a]	C_Seq[e]	II, Jalava *et al.* 1995
533r	TTACCGCGGCTGCTGGCAC	16S rRNA	510 – 528[a]	C_Seq[e]	II, Jalava *et al.* 1995
907r	CCCGTCAATTCCTTTTTGAGTTT	16S rRNA	923-902[a]	C_Seq[e]	II, Jalava *et al.* 1995
23SV_univF_1926	TAAGGTAGCGAAATTCCTTGTCG	23S rRNA	1928-1950[f]	PCR_P[b]	III
Bio_23SV_univR_2259	Bio-CGACCGCCCCAGTCAAACT	23S rRNA	2243–2261[f]	PCR_P[b]	III
23SV_gpos_seq	GGTTACCCGCGACAGGACGG	23S rRNA	2040–2059[f]	Pyroseq[c]	III
23SV_gneg_seq	CCGCGGCAAGACGG	23S rRNA	2046–2059[f]	Pyroseq[c]	III
23SV_Hinf_seq	CCGCGGCTAGACGG	23S rRNA	2046–2059[f]	Pyroseq[c]	III
23SV_myco_seq	TACGYGCGGCGGACGA	23S rRNA	2261-2278[g]	Pyroseq[c]	III
SHV_no_PCR-1	ATGCGTTATATTCGCCTGTG	*shv*	199-218[h]	PCR_P[b]	IV, Nyberg *et al* 2007
SHV_pyro_35_Rbio2	CCGCASAGCASRACTTTA	*shv*	419-402[h]	PCR_P[b]	IV
SHV_cod238_1	CTGGTTTATCGCCGATAAGA	*shv*	870-889[h]	PCR_P[b]	IV
SHV_pyro_238_Rbio	TTGCCAGTGCTCGATCAG	*shv*	1053-1036[h]	PCR_P[b]	IV
SHV_py35F1mod	CGCAGCCGCTTGAGCAAATTA	*shv*	266-286[h]	Pyroseq[c]	IV
SHV_seq_238_240	TATCGCCGATAAGACCGGAG	*shv*	876-895[h]	Pyroseq[c]	IV

[a] According to accession number NC_003098

[b] PCR_P, PCR for pyrosequencing

[c] Pyroseq, pyrosequencing primer

[d] PCR_C, PCR primer for cycle sequencing

[e] C_Seq, primer for cycle sequencing

[f] According to accession number NC_003098

[g] According to accession number NC_008595

[h] According to accession number AF124984

4.5.1 Pyrosequencing dispensation orders (I-IV)

The dispensation orders used in pyrosequencing assays (Table 5) were designed so that at least 30 bp sequences could be generated for all the streptococcal type strains (II). In studies III and IV the dispensation orders were designed to be easily visually interpreted; mutant types could be easily differentiated from wild type sequences.

Table 5. The most often used pyrosequencing nucleotide dispensation orders.

Dispensation order	Description	Study
GCGACTC 10 (ACTG)[a]	v1 region of the 16S rRNA	II
TCTACAGTCGA 10 (CGTA)	v2 region of the 16S rRNA	II
TCGACGAGACATG	position 2058-2059 of the 23S rRNA	III
TACGAGACATG	Confirmation of mutant alleles at the position 2058 of the 23S rRNA	III
TACAGCAGAGTCAGCGAGT	Determination of the sequence corresponding the amino acid position 35 of the SHV gene	IV
ACTGCG5 (ACGT)	Determination of the sequence corresponding the amino acid positions 238-240 of the SHV gene	IV

[a] n (ACGT) signifies cyclic dispensation, nucleotides are added n times in the order A, C, G, T.

4.5.2 Analysis of pyrosequencing results of the streptococcal isolates (I, II)

After a pyrosequencing run, the streptococcal v1 and v2 sequence results were checked manually (I, II). In study I, the sequences of the normal microbiota isolates were manually compared to the sequences of the type strains used in the study. The nonstreptococcal isolates were discriminated from streptococci by their distinct sequence from the beginning and these isolates found their genera by using database searches (Ribosomal Database Project, RDP (Maidak *et al.* 2001)).

In study II, the sequences of the v1 and v2 regions of the type strains listed in Table 2 were determined by pyrosequencing or retrieved from the GenBank (mainly via NCBI, National Center for Biotechnology Information). Thirty-bp sequences of both regions were concatenated and combined in a single Fasta file that could be used as a sequence database in the IdentiFire 1.0.5.0 software (Qiagen). A concatenated sequence of the v1 and v2 sequences of the streptococcal isolates was formed similarly to the type strains, and the sequences of the unknown isolates were compared to the sequences in the database using Identifire software. With this software, it was possible to automatically determine the species whose sequence was closest to the sequence of the unknown isolate. The sequence alignments of the type strains and the unknown isolates were given scores by the software; the score was 100 if the sequences were identical and 96.9 if there was one mismatch in the 60 bp sequence etc. Also gaps generated to the alignment lowered the score, for example, a sequence having 59/60 homologous nucleotides and one gap was given a score of 96.7. For the isolates whose complete 60-base v1+v2 sequence could not be determined, the 30-base sequence of the v1 or v2 region was used. IdentiFire software gave a score of 93.7 for an isolate having one mismatch to the closest type strain sequence in a 30-base sequence. A pyrosequencing

result was regarded as a reliable species-level identification if the pyrosequencing score of an isolate was ≥96.9 to a single species, and results having scores ≥96.9 to more than one species or lower scores were considered group-level identifications, if possible. The isolates identified to *S. mitis* or *S. sanguinis* group are regarded to belong to the combined *S. mitis – S. sanguinis* group in the results and discussion sections of this thesis.

4.5.3 Analysis of pyrosequencing results of the 23S rRNA assay (III)

The agreement of the pyrosequencing results with the known mutation status of the isolates were first estimated visually. Then, the expected and detected numbers of incorporated nucleotides were calculated. In these calculations, the height of the first nucleotide peak that was designed to be negative peak signal control for all samples was used as a signal background check and was subtracted from all successive peak heights. Then, the peak heights were compared to the peak height of a T or an A peak outside the target mutation detection region for which the theoretical number of nucleotide incorporations remained constant irrespective of the mutation status in streptococci. The signals of C and G peaks were compared to the signal of the T peak and the signals of A peaks were compared to the signal of the constant A peak because of the higher signal intensity of the A peaks in pyrosequencing.

Due to different sequence of the 23S rRNA in *C. jejuni, H. influenzae,* and *M. avium* isolates, the signals had to be compared to peaks residing inside the mutation detection region, but remaining constant in case of a mutation at the position 2058.

4.5.4 Analysis of pyrosequencing results of the SHV assay (IV)

The SHV sequences determined by the SHV pyrosequencing assays were determined visually with help of the known SHV sequences (Table 6). The proportions of the SHV sequences were also visually estimated by comparison of heights of constant peaks of each sequence variant.

Table 6. The known SHV sequence variants present at the regions analysed by pyrosequencing.

	Assay	
SHV type	**35**	**238-240**
SHV-1	AACTAAGCGAAAGCCAGCT	CTGGCGAACGGGGTGCGCGCGGG
SHV-1	AACTAAGCGAAAGCCAGCT	CTGGCGAGCGGGGTGCGCGCGGG
SHV-11	AACAAAGCGAAAGCCAGCT	CTGGCGAACGGGGTGCGCGCGGG
SHV-2	AACTAAGCGAAAGCCAGCT	CTAGCGAGCGGGGTGCGCGCGGG
SHV-2a	AACAAAGCGAAAGCCAGCT	CTAGCGAGCGGGGTGCGCGCGGG
SHV-5	AACTAAGCGAAAGCCAGCT	CTAGCAAGCGGGGTGCGCGCGGG
SHV-12	AACAAAGCGAAAGCCAGCT	CTAGCAAGCGGGGTGCGCGCGGG

4.6 Cycle sequencing (II)

For determining the partial 16S rRNA sequence, DNA was amplified using PCR primers fd1mod2 and rp2 (Table 4). The PCRs contained 0.2 pmol/µl primers, 0.030 U/ml AmpliTaq Gold LD DNA polymerase (Applied Biosystems, Foster City, CA, USA), 1x PCR Buffer (Applied Biosystems), 2 mM $MgCl_2$ (Applied Biosystems), and 0.2 mM dNTPs (GE Healthcare). The temperature cycling consisted of initial denaturation for 5 min at 95 °C followed by 40 cycles of denaturation at 94 °C for 15 sec, annealing at 53 °C for 15 sec, and extension for 60 sec at 72 °C. Final extension was 7 min at 72 °C. The PCR products were purified using the High Pure PCR product purification kit (F. Hoffmann-La Roche Ltd, Basel, Switzerland) and sequencing was performed using the Big Dye v 3.0 terminators (Applied Biosystems) and primers 357f, 357r, 533f, 533r, and 907r (Table 4) according to the instructions of the manufacturers.

The sequencing results were analysed using Vector NTI advance 10.1.1 software (Invitrogen Corporation, Carlsbad, CA, USA). In addition, the 16S rDNA sequences of the pharyngeal isolates were subjected to a NCBI BLAST (basic local alignment search tool) (http://www.ncbi.nlm.nih.gov/BLAST/).

4.7 Streptococcal genomic 16S rRNA sequences (unpublished)

On 7th January, 2009, the genomes of 31 streptococcal isolates: *S. equi* ssp. *zooepidemicus* (n=1), *S. agalactiae* (n=3), *S. gordonii* (n=1), *S. mutans* (n=1), *S. pneumoniae* (n=6), *S. pyogenes* (n=13), *S. sanguinis* (n=1), *S. suis* (n=2), and *S. thermophilus* (n=3) have been finished. The 16S rRNA sequences of these isolates were retrieved from the GenBank and the intra-isolate variation of the full-length sequences was checked. In addition, the v1 and v2 sequences found in the genomes were analysed by the Identifire software using the local sequence database of this study.

4.8 VITEK 2 analysis (II)

Fifty-four streptococcal type strains and alpha-haemolytic isolates from blood (n=99) or normal pharyngeal microbiota of elderly persons (n=25) were analysed by the VITEK 2 (bioMérieux) system using the colorimetric GP card and WSVT2-R04.01 software according to the instructions of the manufacturer. In brief, bacterial colonies were suspended in sterile saline corresponding to a turbidity of approximately 0.5 Mc Farland standard. The bacterial suspensions were placed into a cassette of the apparatus and the identification card was placed in a neighbouring slot. The identification cards were thereafter inoculated using a vacuum apparatus of the VITEK 2. The VITEK 2 GP card contained 43 biochemical test substrates that measured carbon source utilization and enzymatic activities.

The results of the various reactions were automatically interpreted to identification results by the VITEK 2 instrument. The species-level identification results were given quality scores by the instrument: Exc (excellent), VG (very good), good, LD (low discrimination). The species results with confidence levels better than LD were

considered as species level identifications except for the "slashline species" (*S. mitis / S. oralis* or *S. bovis / S. lutetiensis*) in this study. If VITEK 2 analysis resulted in low discrimination between two species and an additional Voges-Proskauer (VP) test proposed by the device resulted in the same species or group with the pyrosequencing result, the pyrosequencing and VITEK 2 results were regarded as concordant. In addition, other additional tests were proposed by the VITEK 2 for the resolution of some low-discrimination results. However, these tests were not available in our laboratory and thus, these results could not be confirmed.

5 RESULTS

5.1 Identification of streptococci

5.1.1 Identification of streptococci by pyrosequencing (I, II)

Streptococci were identified by pyrosequencing the variable regions v1 (I) or v1 and v2 of the 16S rRNA (II). The regions used for the identification of streptococci were selected based on the alignment of streptococcal genomic 16S rRNA sequences available in spring 2003; *S. agalactiae* (2603/R and NEM316), *S. mutans* UA159, *S. pneumoniae* (R6 and TIGR4), and *S. pyogene*s (MGAS315, MGAS8232, and M1GAS). In the alignment (Appendix I), the majority of the variation between streptococcal species was found in the region of approximately 100 and 200 bases from the 5' end of the 16S rRNA and these regions were selected as targets for the pyrosequencing assay. In this study, the sequences around positions 100 and 200 are called the v1 and v2 region, respectively. Reverse sequencing primers were used in pyrosequencing the variable regions, and the pyrosequencing reaction of the v1 and v2 regions started at positions 92 and 222 according to the *S. pneumoniae* R6 genomic sequence (accession number NC_003098), respectively.

Differentiation of streptococcal type strains (I, II, unpublished)

In study I, the v1 region was used, and pyrosequencing was considered to be reliable sequencing method in our hands because the v1 sequence of *S. pneumoniae* ATCC 49619 obtained by pyrosequencing was identical to the sequence found in *S. pneumoniae* R6 and TIGR4 genomic sequences. Eight different 30-bp v1 sequences were found among the 12 studied type strains. Thus, all the studied type strains could not be differentiated using the v1 region as *S. anginosus* and *S. gordonii*; *S. constellatus* ssp. *constellatus* and *S. intermedius* as well as *S. sanguinis, S. salivarius,* and *S. vestibularis* had identical v1 sequences, respectively (Table 7).

Table 7. The v1 sequences of the streptococcal type strains analysed in study I.

Type strain	v1 sequence
S. sanguinis DSM 20567	GCAACTCATCCAAGAAGAGCAAGCTCCTCT
S. salivarius DSM 20560	GCAACTCATCCAAGAAGAGCAAGCTCCTCT
S. vestibularis DSM 5636	GCAACTCATCCAAGAAGAGCAAGCTCCTCT
S. parasanguinis DSM 6778	GCAACTCCTCCGCTCGGTGCAAGCACCAAG
S. oralis DSM 20627	GCAACTCATCCGCTCGGTGCAAGCACCAAG
S. cristatus DSM 8249	GCAACTCATCCAGAAGAGCAAGCTCCTCCT
S. pneumoniae ATCC 49619	GCAACTCATCCAGAGAAGCAAGCTCCTCCT
S. mitis DSM 12643	GCAACTCATCCGGAGAAGCAAGCTCCTCCT
S. gordonii DSM 6777	GCAACTCACAGTCTATGGTGTAGCAAGCTA
S. anginosus DSM 20563	GCAACTCACAGTCTATGGTGTAGCAAGCTA
S. intermedius DSM 20573	GCAACTCACAGAATACGGTGTAGTAAACTA
S. constellatus ssp. *constellatus* DSM 20575	GCAACTCACAGAATACGGTGTAGTAAACTA

The 76 type strains included in this study (Table 2) were found to contain 53 different 30-base v1 sequences and 65 different 30-base v2 sequences. For example, *S. pneumoniae, S. pseudopneumoniae* and *S. peroris* had identical sequences in the v1 region whereas *S. mitis* and *S. oralis* type strains could not be differentiated based on the v2 region (Appendix II).

Using the combined v1+v2 sequences (30-base sequences of each region), 68 different sequences were found among the studied streptococcal type strains (Appendix II). Only two of these sequences were found in more than one species, the type strains of the *S. salivarius* group members *S. salivarius, S. vestibularis,* and *S. thermophilus* as well as the *S. bovis* group members *S. bovis* and *S. lutetiensis* were found to contain identical sequences, respectively. In addition, all the species designated as a subspecies of a given species had an identical sequence except for *S. gallolyticus* ssp. *gallolyticus*, *S. gallolyticus* ssp. *pasteurianus* and *S. gallolyticus* ssp. *macedonicus* which all had a distinct sequence. In addition, the type strains of following species pairs that have been designated to belong to the same species had identical sequences: *S. agalactiae* and *S. difficilis, S. gallolyticus* ssp. *gallolyticus* and *S. caprinus, S. infantarius* ssp. *coli,* and *S. lutetiensis* as well as *S. gallolyticus* ssp. *macedonicus* and *S. waius.*

The v1+v2 sequence of *S. pneumoniae* differed by two nucleotides from the sequences of *S. pseudopneumoniae* and *S. mitis*. Interestingly, the v1+v2 sequence of the *S. oralis* type strain resembled most the *S. gallinaceus* and *S. parasanguinis* type strain sequences with four and five differences, respectively. *S. pyogenes* strains had a distinguishable v1+v2 sequence that differed from all other streptococcal type strain sequences by at least four nucleotides (Appendix II).

Identification of normal microbiota isolates (I, II, unpublished)

In study I, the normal microbiota isolates (n=172) were identified by pyrosequencing the v1 region. Nine of the isolates were identified as *Enterococcus* and two as *Granulicatella* species resulting in 161 streptococcal isolates. The streptococcal isolates were identified as *S. mitis* (n=43), *S. oralis* (n=25), *S. parasanguinis* (n=23), *S. cristatus* (n=22), and *S. pneumoniae* (n=18) which can all be regarded members of the combined *S. mitis- S. sanguinis* group. In addition, even the streptococcal group of 25 isolates could not be reliably determined as they had the v1 sequence shared by *S. gordonii* and *S. anginosus* (n=9) or by *S. sanguinis, S. salivarius*, and *S. vestibularis* (n=16) type strains (Table 7). In addition, five isolates could be identified only as streptococci due to their heterogeneous 16S rRNA sequences.

In study I, relatively short sequences, as few as 19 nucleotides generated with the primer StepV1rev corresponding to a 16-base sequence obtained with the primer StrepV1RevV2 used in study II were accepted for inclusion to a certain species since the variation seen in the studied type strains was limited and short sequences seemed to suffice for species determination. The result obtained using the v1 region (I) was usually indicative of the result of the combined v1 and v2 region (II) as most of the isolates (24/34; 70.5%) had sequences that indicated the species to belong to the same streptococcal group using both approaches. Six (17.6%) isolates did not reach the

group level agreement because their v1 sequence was shared by streptococcal type strains belonging to different groups (Table 8).

Of the 34 pharyngeal isolates analysed in study II, 11 (32.4%) contained a combined v1 and v2 sequence having a score ≥96.9 to one species and could be reliably identified as *S. mitis* (n=6), *S. gordonii* (n=2), *S. infantis* (n=1), *S. parasanguinis* (n=1), or *S. sanguinis* (n=1) using pyrosequencing (Table 8). Five isolates having a score ≥96.9 to at least two species were identified as *S. salivarius* group members, and two isolates having a sequence differing by one nucleotide from the type strain sequences of *S. pseudopneumoniae, S. pneumoniae,* and *S. mitis* were identified as *S. mitis* - *S. sanguinis* group members. The rest of the isolates had lower pyrosequencing scores and could be identified only to the group level. Even the group-level identification of isolates having very low scores, even below 90, could be considered as only probable. The 16S rRNA sequence of the v1 region was found to be heterogeneous in six (17.6%) isolates, and these isolates could be identified to a probable streptococcal group based on the sequence of the v2 region (Table 8).

Based on the analysis of the combined sequence of the v1 and v2 regions, the vast majority of the isolates (n=30, 88.2%) belonged to the combined *S. mitis* - *S. sanguinis* group, and two (5.9%) to *S. salivarius* group. In addition, one isolate could not be identified even to the group level because its v1+v2 sequence was closest to the sequence of the ungrouped *S. gallinaceus* type strain.

The v2 sequence of 17/18 isolates identified as *S. pneumoniae* in study I was determined, and all these isolates were identified as members of the *S. mitis* - *S. sanguinis* group with the combined v1+v2 sequence (Table 8), and four isolates were identified as *S. pseudopneumoniae* (n=2) or *S. peroris* (n=2).

The 16S rRNA sequence of nine isolates having a sequence differing from all the streptococcal type strain sequences was determined by cycle sequencing (GenBank accession numbers EF151144-EF151152). The sequences determined by cycle sequencing were in agreement with the sequences determined by pyrosequencing, and the BLAST search results were in accordance to the results obtained with the Identifire software using the local sequence database.

Table 8. Comparison of the pyrosequencing results of the v1 region (I) and the combined v1 and v2 region (II, partly unpublished) of the streptococcal normal microbiota isolates included in this study.

n[a]	v1 result	n[b]	v1+v2 result[c]	Score
10	*S. mitis*	1	***S. australis***	87.4
		6	***S. mitis***	≥96.9
		3	**v1 heterogeneous, v2 *S. mitis* - *S. sanguinis* group member**[d]	
3	*S. oralis*	2	*S. parasanguinis*	≤93.7
		1	***S. oralis***	90.5
7	*S. parasanguinis*	1	***S. parasanguinis***	100
		4	***S. parasanguinis***	≤93.7
		1	*S. gallinaceus*	90.5
		1	**v1 heterogeneous, v2 *S. parasanguinis***	
4	*S. cristatus*	3	***S. infantis***	84.2-96.9
		1	***S. cristatus, S. infantis,* or *S. peroris***	93.7
17[e]	*S. pneumoniae*	2	***S. pseudopneumoniae***	100
		2	***S. peroris***	≥96.9
		2	***S. peroris***	≤93.7
		11	***S. pneumoniae, S. pseudopneumoniae,* or *S. mitis***	≤96.9
2	*S. gordonii* or *S. anginosus*	2	*S. gordonii*[f]	96.9
4	*S. sanguinis,*	2	*S. sanguinis*[f]	90.5, 96.9
	S. salivarius, or *S. vestibularis*	2	*S. salivarius / S. vestibularis*[f]	≥96.9
2	*Streptococcus* sp.	1	v1 heterogeneous, v2 *S. mitis, S. oralis,* or *S. infantis*[g]	93.7 (v2)
		1	v1 heterogeneous, v2 *S. parasanguinis*[g]	88.8 (v2)

[a] The number of isolates with a given v1 pyrosequencing result (I).
[b] The number of isolates with a given v1+v2 pyrosequencing result (II).
[c] The v1+v2 results are bold and underlined if the v1 and v1+v2 results coincided at the species level and bold if they coincided at the group level. *S. pneumoniae* is regarded as a member of the *S. mitis*-*S. sanguinis* group in this table.
[d] One isolate had the v2 sequence of *S. australis*, one of *S. mitis* or *S. oralis* and one had a score 93.7 to *S. mitis, S. oralis*, and *S. infantis* type strains.
[e] Two of the isolates identified as *S. pneumoniae* in study I were included in study II and were identified as *S. pneumoniae, S. pseudopneumoniae*, or *S. mitis*. *S. pneumoniae* was considered as a member of the *S. mitis* group in this table.
[f] The species of the v1 result belong to different streptococcal groups.

Identification of clinical isolates (II)

The invasive *S. pneumoniae* isolates (n=17) included in this study were found to have the *S. pneumoniae* –specific v1 and v2 sequences by pyrosequencing.

Only 37 of the blood culture isolates (n=99) had a pyrosequencing score ≥96.9 to a single streptococcal type strain sequence (Table 2, Appendix II) and could be unequivocally designated to these species, most commonly to *S. anginosus* (n=10), *S. sanguinis* (n=9), *S. intermedius* (n=6), or *S. mitis* (n=4) (Table 9).

In addition, 59 (59.6%) of the isolates could be identified to the group level. The most common sequence among these isolates (n=24) was a sequence differing by one nucleotide from the type strain sequences of the *S. mitis* group members *S. pneumoniae*, *S. pseudopneumoniae,* and *S. mitis* and thus these isolates were identified to the *S. mitis* - *S. sanguinis* group (Table 9).

In total, 61 isolates were identified to the *S. mitis* - *S. sanguinis*, 21 to the *S. anginosus*, 7 to the *S. bovis*, and 7 to the *S. salivarius* group. The streptococcal group of three isolates could not be determined due to heterogeneous sequences detected in the pyrograms. These included one isolate having a heterogeneous sequence in both regions and two isolates having heterogeneous v2 sequence and sequence of *S. oralis* or ungrouped *S. gallinaceus* in the v1 region.

Table 9. Pyrosequencing results of the combined v1 and v2 region sequence of the 99 blood culture isolates.

n	Pyrosequencing result	Score	Streptococcal group
Species-level identification[a]			
10	*S. anginosus*		*S. anginosus*
6	*S. intermedius*		*S. anginosus*
3	*S. constellatus*[b]		*S. anginosus*
9	*S. sanguinis*		*S. mitis* - *S. sanguinis*
4	*S. mitis*		*S. mitis* - *S. sanguinis*
2	*S. gallolyticus* ssp. *gallolyticus*		*S. bovis*
1	*S. oralis*		*S. mitis* - *S. sanguinis*
1	*S. gordonii*		*S. mitis* - *S. sanguinis*
1	*S. gallolyticus* ssp. *pasteurianus*		*S. bovis*
Group-level identification[b]			
24	*S. pseudopneumoniae*, *S. pneumoniae* or *S. mitis*	96.9	*S. mitis* - *S. sanguinis*
6	*S. infantis*	93.7	*S. mitis* - *S. sanguinis*
2	*S. oralis*	≤93.7	*S. mitis* - *S. sanguinis*
1	*S. peroris*	93.7	*S. mitis* - *S. sanguinis*
9	*S. sanguinis*	≤93.7	*S. mitis* - *S. sanguinis*
1	*S. gordonii*	87.4	*S. mitis* - *S. sanguinis*
3	v1 heterogeneous, v2 *S. mitis* or *S. oralis*	100 (v2)	*S. mitis* - *S. sanguinis*
4	*S. bovis* or *S. lutetiensis*	≥96.9	*S. bovis*
2	*S. constellatus* [b] or *S. intermedius*	96.9	*S. anginosus*
7	*S. salivarius* or *S. vestibularis*	≥96.9	*S. salivarius*
Genus-level identification			
2	v2 heterogeneous, v1 *S. oralis* or *S. gallinaceus*	100 (v1)	unknown
1	Streptococcus	-	-

[a] Score ≥ 96.9 to one species

[b] Isolates were identified to the streptococcal group based on the species result irrespective of the pyrosequencing score. If the pyrosequencing had the same score to more than one species, group-level identification was reached if the species belonged to the same streptococcal group.

Sequence variations in the streptococcal 16S rRNA sequences (I, II)

Surprisingly high number of streptococcal isolates was found to have v1 or v2 sequences differing from all the studied streptococcal type strain sequences. The degree of sequence variation detected by pyrosequencing was found to differ substantially between sample types; sequence variation was not found among studied invasive *S. pneumoniae* isolates, whereas the proportion of strains that did not have an identical sequence among the streptococcal type strain sequences was high among the pharyngeal normal microbiota isolates (Table 10).

Table 10. Pyrosequencing scores according to the different isolate types.

	Score[a]		
Isolate type	**≥96.9 to one species**	**≥96.9 to more than one species**	**<96.9 to all species**
S. pneumoniae (n=17)	17 (100 %)	0	0
Blood culture isolates (n=99)	37 (39.4 %)	37 (37.4 %)	25 (25.3%)
Pharyngeal isolates (n=34)	11 (32.4 %)	3 (17.6 %)	20 (70.6 %)

[a] Number (%) of isolates with a given pyrosequencing score

Apparently heterogeneous (Figure 6) 16S rRNA sequences were found in 6/34 (17.6%) normal microbiota and 6/99 (6.1 %) blood culture isolates. Most of the heterogeneous sequences (9/12) were found in the v1 region. It is however not possible to be certain that these were indeed heterogeneous sequences in single isolates and not double isolates. The possibly contaminated isolates were however tried to be separated by subculturing, and even after several cultivation rounds, the isolates had similar phenotype and the same, heterogeneous pyrosequencing pattern.

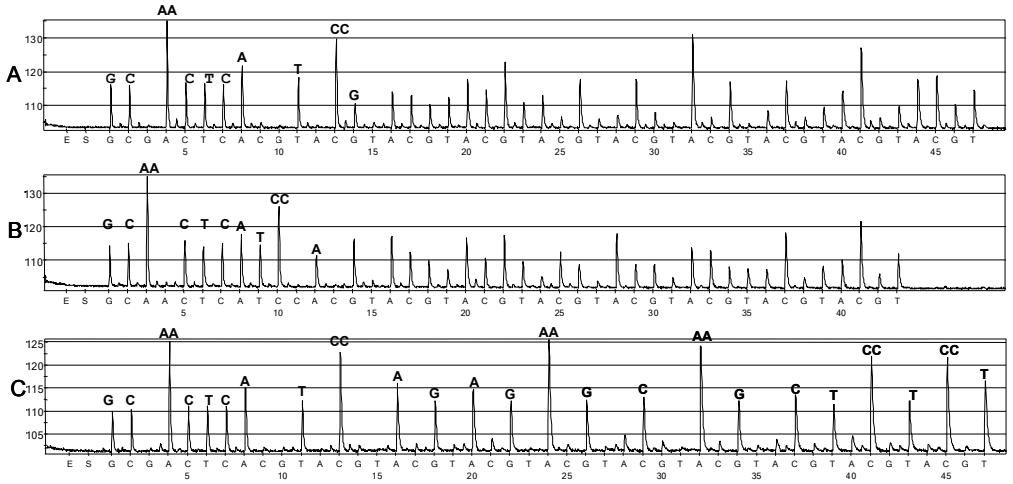


Figure 6. Pyrograms indicating heterogeneous v1 sequences (A and B) of a pharyngeal streptococcal isolate and a pyrogram of the v1 region of *S. pneumoniae* (C). Because the pyrograms A and B are from the same isolate, the isolate probably contains v1 sequences beginning by GCAACTCATCCG and GCAACTCATCCA

Streptococcal genomic 16S rRNA sequences (unpublished)

In order to check the presence of heterogeneous sequences within streptococci, 16S rRNA sequences of the published streptococcal genomes were analysed. The 16S rRNA sequences found within streptococcal genomes except *S. gordonii* str. Challis substr. CH1 and *S. thermophilus* LMD-9 were found to be identical to each other.

When the genomic sequences were compared to the combined v1 and v2 sequences of the streptococcal type strains, only the sequences of *S. gordonii* str. Challis substr. CH1, *S. thermophilus* LMD-9, and *S. equi* ssp. *zooepidemicus* were found to differ from the sequence of the respective type strain. The four 16S rRNA sequences of *S. gordonii* str. Challis substr. CH1 differed from the *S. gordonii* type strain sequence in the v1 and v2 region, the pyrosequencing scores of its two alleles compared to the *S. gordonii* type strain were 87.4 and 90.5. *S. thermophilus* LMD-9 contained six 16S rRNA copies, three of which had an identical v1+v2 sequence to the *S. thermophilus* type strain and three had a sequence that differed by three nucleotides from the *S. thermophilus* type strain. Furthermore, the v1 and v2 sequence found in all the 16S rRNA copies of the genome of *S. equi* ssp. *zooepidemicus* MGCS10565 was not closely related to any type strain sequence; it was closest to *S. suis* or *S. equi* ssp. *ruminatorum* type strain sequence (score 81 to both).

5.1.2 Identification of streptococci by VITEK 2 (II)

Streptococcal type strains (II)

Fifty-seven streptococcal type strains were analysed by VITEK 2 (II), and 29 (50.9 %) of them were correctly identified to the species level and two additional type strains were identified to the correct species with low differentiation with some other species. In addition, *S. acidominimus* type stain was correctly identified as unidentified organism because this species was not included in the species database of VITEK 2. Furthermore, fifteen type strains were identified to the correct streptococcal group by VITEK 2, and ten type strains could not be identified even to the correct group by the VITEK 2.

Majority of the streptococcal type strains that could not be identified to the correct species with VITEK 2 (20/26, 76.9%) and all the type strains that could not be identified to the group level belonged to a species that was not included in the VITEK database containing mainly well-established species relevant to human infections. Furthermore, the type strains that could not be identified correctly were often identified to a wrong species even with an excellent confidence level.

Of the type strains regarded as viridans streptococci or non-beta-haemolytic streptococci (n=40), 20 (50.0%) and 33 (82.5%) could be identified to the correct species and group by VITEK 2, respectively. In addition, two type strains, the *S. parasanguinis* and *S. caprinus* type strains, were identified to the correct species and another streptococcal species with low discrimination. Six alpha-haemolytic type strains (13.6%) could not be identified to the correct group.

Alpha-haemolytic isolates (II)

Majority of the 99 blood culture isolates (n=85, 85.9%) were identified to the species level, three isolates to the group level, two isolates could be identified as streptococci, and nine isolates (9.1%) could not be identified using VITEK2 (Table 11).

The majority of the isolates (59/99, 59.6%) were identified to the combined *S. mitis – S. sanguinis* group, and the most common identification result was the "slashline species" *S. mitis / S. oralis* (n=45, 45.5%). *S. sanguinis* (n=9), *S. gordonii* (n=2), and *S. parasanguinis* (n=1) isolates were identified to the species level. Eighteen isolates (18.2%) were identified to the *S. anginosus* group, sixteen of which were identified to the species-level. Seven (8.1%) isolates were identified to a species belonging to the *S. bovis* group by VITEK 2 (Table 11).

Table 11. VITEK 2 results of the blood culture isolates.

n	VITEK 2 result	Confidence level (n) [a,b]	Streptococcal group
Species-level identification			
9	*S. anginosus*	Exc (5), LD (4)	*S. anginosus*
2	*S. constellatus* ssp. *constellatus*	Exc, LD	*S. anginosus*
1	*S. constellatus* ssp. *pharyngis*	Good	*S. anginosus*
4	*S. intermedius*	Exc	*S. anginosus*
2	*S. gallolyticus*	Exc	*S. bovis*
4	*S. infantarius*	Exc (2), VG (2)	*S. bovis*
1	*S. pasteurianus*	Exc	*S. bovis*
2	*S. gordonii*	Exc	*S. mitis - S. sanguinis*
1	*S. parasanguinis*	Vg	*S. mitis - S. sanguinis*
9	*S. sanguinis*	Exc (8), Good (1)	*S. sanguinis*
5	*S. salivarius*	Exc	*S. salivarius*
Group-level identification			
45	*S. mitis / S. oralis*	Exc (22), VG (12), Good (8), Acc (1), LD (2)	*S. mitis - S. sanguinis*
1	*S. constellatus* ssp. *constellatus* or *S. anginosus*	LD	*S. anginosus*
1	*S. intermedius* or *S. constellatus*	LD	*S. anginosus*
1	*S. parasanguinis* or *S. mitis / S. oralis*	LD	*S. mitis - S. sanguinis*
Genus-level identification			
1	*S. gallolyticus, S. bovis / S. lutetiensis,* or *S. hyointestinalis*	LD	*S. bovis* or ungroupable
1	*S. pluranimalium* or *S. mitis / S. oralis*	LD	*S. mitis - S. sanguinis* or ungroupable
Unidentified			
9	inconclusive identification or unidentified organism		-

[a] If required, the number of isolates with a given confidence level is in parentheses

[b] For the isolates having confidence level LD, a supplemental VP test was performed, and the result presented in this table is the result obtained with the additional VP test

Twenty-five isolates from the normal microbiota were analysed by VITEK 2 (II), and all but one isolate could be identified to the species level (Table 12). The majority of the pharyngeal isolates (n=20) belonged to the *S. mitis* group based on the VITEK 2 analysis. In addition, three *S. anginosus* group members and one *S. salivarius* isolate were found. The confidence levels of the results of the pharyngeal isolates were high: the results of 22 isolates were obtained with at least confidence level good.

Table 12. VITEK 2 results of the pharyngeal isolates.

n	VITEK 2 result	Confidence level (n)	Streptococcal group
10	*S. mitis / S. oralis*	Exc (4), good (1), VG (4), acc (1)	*S. mitis*
1	*S. gordonii*	Exc	*S. mitis*
9	*S. parasanguinis*	Exc (5), vg (3), good (1)	*S. mitis*
1	*S. constellatus* ssp. *pharyngis*	good	*S. anginosus*
1	*S. intermedius*	Exc	*S. anginosus*
2	*S. salivarius*	Exc, acc	*S. salivarius*
1	inconclusive identification	-	-

5.1.3 Comparison of pyrosequencing and VITEK 2 results (II)

The pyrosequencing and VITEK 2 results of an isolate were considered concordant at the species level if the pyrosequencing score was ≥96.9 to a single species and VITEK resulted in the same species with pyrosequencing. When the results of the 124 isolates analysed by both methods were compared, species-level agreement was found in 30 (24.2%) isolates (Table 13). Group-level agreement was reached for 74 (59.7%) isolates and for 10 isolates, genus-level agreement between the two methods was found as these isolates were identified to different streptococcal groups using the two methods or to an ungrouped streptococcal species by other method. Ten (8.0%) isolates identified as various streptocci by pyrosequencing could not be identified using VITEK2 and resulted in contradictory identification results with the two methods. Thus, for 104 isolates (81.9%), an identification result concordant at least at the group level was obtained with the two methods.

Table 13. Comparison of the pyrosequencing and VITEK 2 results of the blood culture and pharyngeal isolates analysed by both methods.

n	Pyrosequencing result	Vitek 2 result
Species-level agreement		
9	*S. sanguinis*	*S. sanguinis*
2	*S. gordonii*	*S. gordonii*
1	*S. parasanguinis*	*S. parasanguinis*
9	*S. anginosus*	*S. anginosus*
5	*S. intermedius*	*S. intermedius*
1	*S. constellatus*[a]	*S. constellatus* ssp. *constellatus*
2	*S. gallolyticus* ssp. *gallolyticus*	*S. gallolyticus*
1	*S. gallolyticus* ssp. *pasteurianus*	*S. pasteurianus*
Group-level agreement		
7	*S. mitis*	*S. mitis* or *S. oralis*
1	*S. oralis*	*S. mitis* or *S. oralis*
45	*S. mitis – S. sanguinis* group member	*S. mitis* or *S. oralis*
1	*S. mitis – S. sanguinis* group member	*S. parasanguinis*
5	*S. parasanguinis* (93.7) [b]	*S. parasanguinis*
2	*S. parasanguinis* (v2)	*S. parasanguinis*
1	*S. gordonii* (87.4) [b]	*S. gordonii*
1	*S. intermedius*	*S. constellatus* ssp. *pharyngis*
1	*S. anginosus* group member	*S. constellatus* ssp. *constellatus*
1	*S. anginosus* group member	*S. anginosus* group member
3	*S. bovis* or *S. lutetiensis*	*S. infantarius*
6	*S. salivarius* or *S. vestibularis*	*S. salivarius*
Genus-level agreement		
1	*S. mitis*	*S. intermedius*
1	*S. mitis – S. sanguinis* group member[c]	*S. pluranimalium* or *S. mitis* or *S. oralis*
1	*S. mitis – S. sanguinis* group member[c]	*S. constellatus* ssp. *pharyngis*
1	*S. mitis – S. sanguinis* group member[c]	*S. salivarius*
1	*S. bovis* or *S. lutetiensis*	*S. bovis* group member or *S. hyointestinalis*
1	*S. oralis* or *S. gallinaceus* (v1 score 100) [c]	*S. parasanguinis*
1	*S. oralis* or *S. gallinaceus* (v1 score 100) [c]	*S. constellatus* ssp. *constellatus*
1	*S. gallinaceus* (score 91)	*S. mitis* or *S. oralis*
1	*S. salivarius* or *S. vestibularis*	*S. infantarius*
1	*Streptococcus* (v1+v2 heterogeneous)	*S. mitis* or *S. oralis*
Controversial result		
1	*S. salivarius* or *S. vestibularis*	unidentified organism
5	*S. mitis – S. sanguinis* group member	unidentified organism
1	*S. mitis – S. sanguinis* group member	inconclusive identification
1	*S. anginosus*	inconclusive identification
2	*S. constellatus*	inconclusive identification

[a] *S. constellatus* ssp. *constellatus* or *S. constellatus* ssp. *pharyngis*

[b] These isolates are considered to have only group level agreement as the pyrosequencing results had a score ≤93.7 or heterogeneous sequence in the v1 region.

[c] Heterogeneous sequence in either v1 or v2 region.

5.2 Detection of macrolide resistance mutations in the 23S rRNA (III)

The method developed for the detection of mutations causing macrolide resistance at the positions 2058-2059 of the 23S rRNA was tested for *S. pneumoniae*, *S. pyogenes, Mycobacterium avium, Campylobacter jejuni,* and *Haemophilus influenzae* isolates and the PCR primers of the assay amplified DNA from all the tested bacterial isolates.

With the appropriate sequencing primer, distinctive pyrograms were obtained for wild-type and resistant isolates (Figure 7) after template preparation and an approximately 15 min pyrosequencing run.

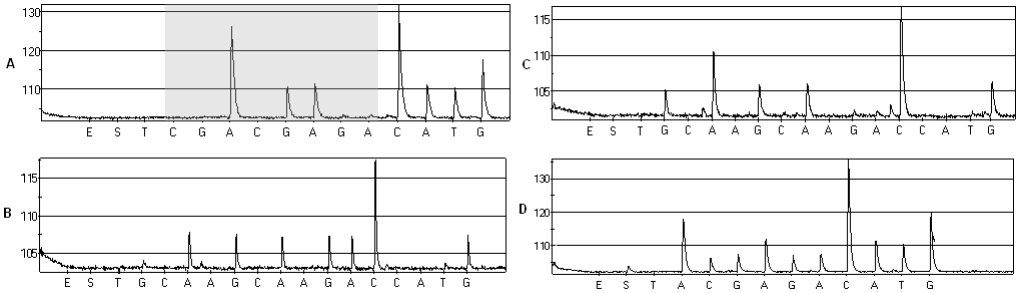


Figure 7. Pyrograms obtained with the 23S rRNA assay. A) *S. pneumoniae* ATCC 700294, wild type. The mutation detection region of the pyrogram is indicated by grey background. B) *Campylobacter jejuni* 62, mutation A2059G in all 23S rRNA copies C) *Haemophilus influenzae* 286, A2058G mutation in all 23S rRNA copies, D) *S. pneumoniae* r506, A2059C mutation in two of four 23S rRNA alleles.

The pyrosequencing assay was also found to be quantitative as the peak heights followed well the known number of wild type and mutated 23S rRNA copies in the isolates (Table 3). Of the previously unknown strains, one copy containing the A2059C mutation could be detected among three wild type copies in *S. pneumoniae* r581 and *S. pneumoniae* 01-41 was found to carry two wild type and two A2058G alleles. All alleles were found to contain a mutation at the position 2058 or 2059 of the 23S rRNA in the studied erythromycin-resistant *H. influenzae* and *C. jejuni* isolates. The studied azithromycin-resistant *M. avium* isolate contained the A2058G mutation whereas the susceptible isolate contained wild-type 23S rRNA.

The importance of the design of the dispensation order was clearly seen as one wild type copy among five mutated 23S rRNA copies of *S. pyogenes* could not be detected using the standard dispensation order whereas it could be detected with a dispensation order in which a specific peak was designed to appear for the wild-type copy (Figure 8).

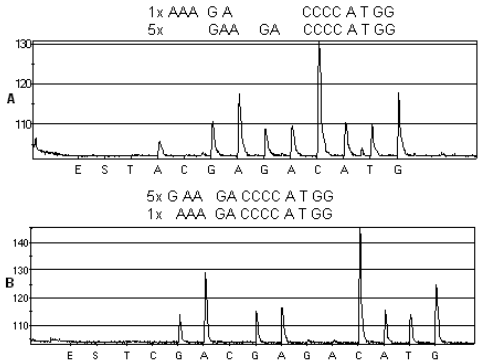


Figure 8. Pyrograms obtained with A) the dispensation order designed for detecting mutated alleles at the position 2058 of the 23S rRNA B) the normal dispensation order with the isolate *S. pyogenes* NI4277 containing one wild type and five 23S rRNA copies having the A2058G mutation.

5.3 Typing of SHV by pyrosequencing (IV)

The primers used for the SHV assay successfully amplified DNA from all the studied SHV-positive *K. pneumoniae* and *E. coli* isolates and the SHV sequence of the Ambler amino acid positions 35, 238 and 240 could be determined visually from the obtained pyrograms. The pyrosequencing assays of this study are not able to determine the exact SHV type as only the positions 35, 238 and 240 are studied, and the results are referred to as "SHV-1 or SHV-5 pyrosequencing type", for example.

The pyrosequencing result of an isolate was combined from the pyrograms of the assays covering position 35 and positions 238-240, and if two or more sequences were detected, the most prevalent sequence was determined from the quantitative peak heights (Figure 9).

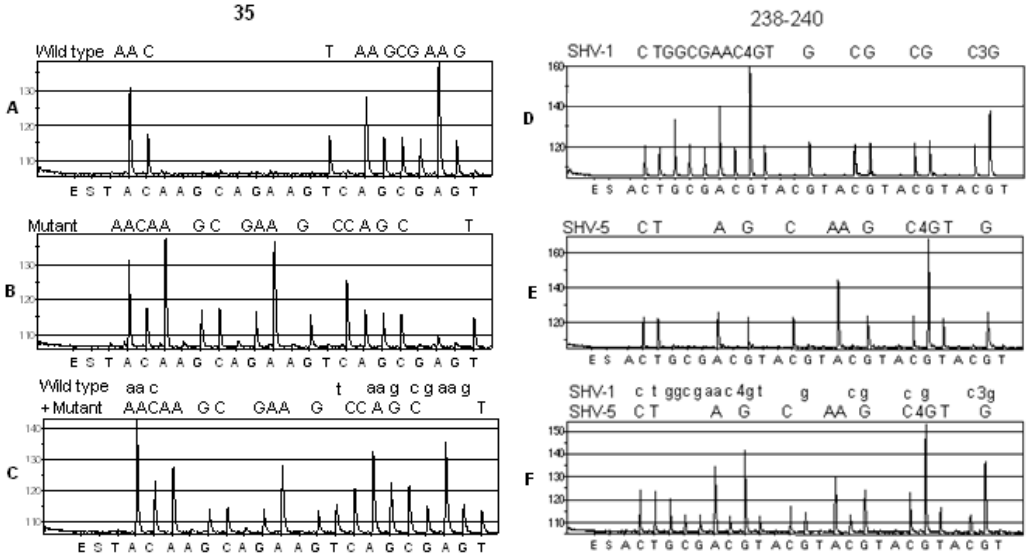


Figure 9. Sample pyrograms of the SHV pyrosequencing assays. A-C) pyrograms of the position 35 assay, of a sample with a wild type SHV sequence (SHV-1 or SHV-11) (A), mutant sequence (B), and wild type and mutant sequences (C). D-F) Pyrograms of the position 238-240 assay of a sample containing SHV-1 (D), SHV-5 (E), and both SHV-5 and SHV-1 pyrosequencing types (F).

In the studied *E. coli* isolates (n=10), the SHV sequences of the positions 35 and 238-240 determined by pyrosequencing were in agreement with the SHV types determined by cycle sequencing SHV-2 (n=2), SHV-2a (n=1), SHV-5 (n=2), SHV-11 (n=1), and SHV-12 (n=4).

The SHV type of 40 *K. pneumoniae* isolates that could not be determined by cycle sequencing due to the heterogeneous peaks in the cycle sequencing electropherograms (Figure 10) could be defined by pyrosequencing. All these isolates were found to contain two SHV copies differing at the position 35, 238 and/or 240 (Figure 9).

In the 66 *K. pneumoniae* isolates for which an SHV type had been determined earlier by cycle sequencing, the pyrosequencing results were in agreement with the cycle sequencing result in 34/66 (51.5 %) isolates. The SHV types found in these isolates were SHV-1 (n=16), SHV-11 (n=14), SHV-12 (n=3), SHV-2 (n=1), and SHV-12+SHV-1 (n=1). However, in 31 of these isolates, more than one SHV copy was detected by pyrosequencing indicating possibility of presence of a plasmid and a chromosomal SHV copy.

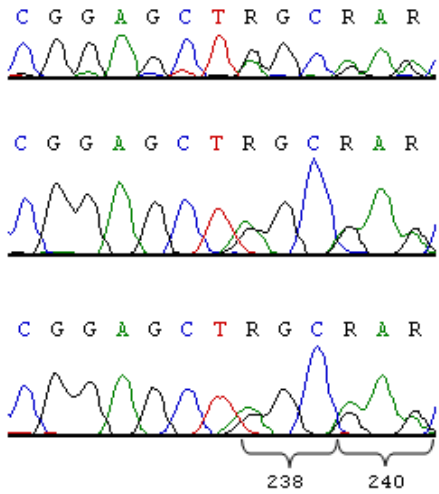


Figure 10. Example of cycle (Sanger) sequencing electropherograms with ambiguous sequence at the positions 238 and 240 of the SHV molecule. In the figure, three nucleotide positions are ambiguous: the first nucleotide of the codon 238 and the first and third nucleotides of the codon 240.

Usually the pyrosequencing results were in agreement with the cycle sequencing results as the ESBL sequence that had been found by cycle sequencing was also detected by pyrosequencing. The pyrosequencing result was not in agreement with the cycle sequencing result in 4/66 (6.1 %) isolates, as the two types detected by pyrosequencing differed from the original type. The discrepant results of these isolates were due to the heterogeneous sequences in the positions that were pyrosequenced in this study; for example one isolate was identified as SHV-2 by cycle sequencing and was in this study found to carry SHV-5 and SHV-11 pyrosequencing types. Another isolate was found to contain three SHV copies by pyrosequencing: two different SHV-1 copies, and in addition, the SHV-2 gene that had already been determined by cycle sequencing (Figure 11).

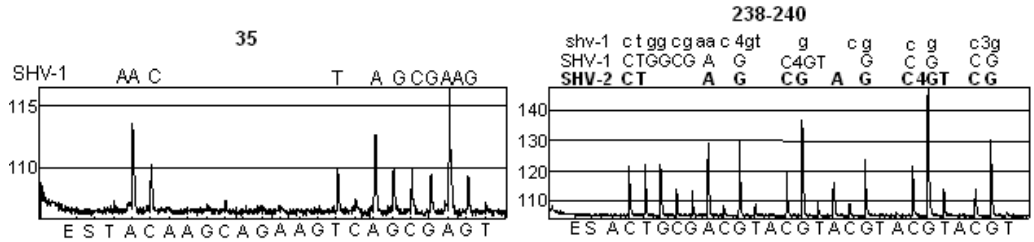


Figure 11. Pyrograms indicating three different SHV sequences in an isolate. The sequences differ at the positions 238-240 but all are wild type at the position 35.

Consequently, more than one SHV copy was present in 72 of 106 (67.9%) studied Finnish *K. pneumoniae* isolates, and 66 (91.7 %) of these were found to contain an ESBL copy by pyrosequencing and in six (8.3 %) isolates two copies of SHV-1 or SHV-11 were found. The most common SHV combination was SHV12 + SHV-11 which was found in 35 (33.0 %) of the studied *K. pneumoniae* isolates.

6 DISCUSSION

6.1 Identification of streptococci by pyrosequencing (I, II)

6.1.1 16S rRNA as the target

The 16S rRNA was chosen as the target for the streptococcal identification method because of its several advantages. 16S rRNA (gene) sequences are well-studied and are extensively used as a target for bacterial species identification. Sequencing-based identification is most advantageous for the identification of fastidious or uncultivable bacteria or isolates having atypical phenotypic characters (Woo *et al.* 2008). In the 16S rRNA, there are constant regions and variable regions that have different degree of variability in different bacteria. Consequently, some regions are more useful for the identification of certain bacteria than other regions (Clarridge 2004). Furthermore, silent mutations do not exist in the ribosomal genes (Clarridge 2004) and the 16S rRNA has been shown to act as a chronometer; the more divergent the 16S rRNA sequences between species are, the longer time has passed from their divergence (Woese 1987). Due to the wide use of this region, there are a large number of 16S rRNA sequences available in the public databases to which unknown sequences can be compared. However, most of the sequencing data found in public databases is not reliable (Hoshino *et al.* 2005; Hugenholtz and Huber 2003; Yarza *et al.* 2008).

The 16S rRNA has however been claimed to be too conserved in the streptococci to be used for their identification and several other genes as *ddl* (D-alanine:D-alanine ligase), 23S rRNA, *rnpB*, *tuf*, and *sodA* have been proposed for the identification of streptococci (Hoshino *et al.* 2005; Picard *et al.* 2004). In recent studies, it has also been suggested that identification of streptococci should not be based on a single target, as the streptococcal isolates cluster differently using different gene targets due to horizontal gene transfer that has been detected at least among the members of the *S. anginosus*, *S. mitis,* and *S. salivarius* groups (Delorme *et al.* 2007; Hoshino *et al.* 2005; Jacobs *et al.* 2000; Kilian *et al.* 2008). Due to the very common gene exchange among streptococcal species, any sequencing result should be considered as only indication of belonging to certain taxonomical cluster.

6.1.2 Identification of streptococci by pyrosequencing the 16S rRNA (I, II)

To the pyrosequencing assay of this study, the most variable regions in streptococcal 16S rRNA, v1 and v2, were chosen. Only sequences of streptococcal type strains were included to the sequence database of the assay. This is because only the type strains are by definition known representatives of the species. There were various versions available of the type strain sequences in the GenBank at the beginning of the study. To obtain a reliable sequence database, the v1 and v2 sequences of most of the streptococcal type strains were analysed by pyrosequencing in study II. Afterwards however, a sequence that was identical to the sequence determined by pyrosequencing could be found among the various sequence versions available in the GenBank for most type strains. The finding of the correct type strain sequence would today be a simpler task since the 16S rRNA sequences of all bacterial type strains have been collected to a database, Eztaxon (Chun *et al.* 2007).

In study I it was noticed that the v1 region did not differentiate even all the type strains included in the study as several species had identical sequences. In the earlier use of pyrosequencing for microbial and viral analysis, however, only few pyrosequencing assays for the identification of viruses (Gharizadeh *et al.* 2001) or bacteria (Jonasson *et al.* 2002) had been described, and they mainly relied on a single region. However, *H. pylori* had been identified and typed by pyrosequencing the v1 and v3 regions of the 16S rRNA (Monstein *et al.* 2001). Later on, the results of other studies have been very similar to the results of this study; for example several mycobacterial species could not be differentiated from each other based on a single pyrosequencing reaction of the variable region of the 16S rRNA gene (Kirschner and Bottger 1998; Tuohy *et al.* 2005). This was however not considered a problem as the major pathogens, *M. tuberculosis* and *M. avium-intracellulare* complexes, could be differentiated from the other mycobacterial species.

In study II, the combined sequence of the v1 and v2 regions was found to differentiate the streptococcal type strains well, as the few type strains that had identical v1 and v2 sequences were either very closely related species or species currently considered to belong to a single species. The differences between the closest differentiated type strain sequences were also minimal, only one nucleotide in the members of the *S. bovis* group and two nucleotides in the *S. mitis* group and *S. pneumoniae*. Identification of streptococci using the *rnpB* also required pyrosequencing of two regions to achieve adequate discrimination of streptococcal species, and *S. anginosus* and *S. constellatus* could not be differentiated from each other even using the two regions (Innings *et al.* 2005).

6.1.3 Sequence variation in the streptococcal 16S rRNA (I, II)

S. pneumoniae ATCC 49619, DSM 11867, and invasive *S. pneumoniae* isolates (n=17) were found to have identical v1 and v2 sequences to the *S. pneumoniae* type strain in this study. In addition, most of the analysed genomic sequences contained identical sequence to their respective type strains. However, only minority of the studied alpha-haemolytic isolates, (43/124, 34.7%) had a v1+v2 sequence of the 16S rRNA that was identical to a streptococcal type strain sequence in this study. The degree of variation detected in the 16S rRNA sequences of this study was rather surprising, especially when compared to the results obtained by pyrosequencing of the *rnpB* gene. By *rnpB,* 85% of the 113 studied isolates had a sequence that was identical to a sequence in the *rnpB* sequence database. However, the *rnpB* database contained also other than type strains and the majority of the studied isolates were not alpha-haemolytic (Innings *et al.* 2005).

It is however quite common that the sequences found in bacterial isolates differ from known sequences (Suau *et al.* 1999) and sequencing results have even led to reclassification of genera (Collins *et al.* 1994). The phenomenon has also been detected in other studies on VGS: some isolates could not be reliably assigned to a species because the 16S rRNA, *rpoB,* and/or *tuf* sequences of an isolate had similar degrees of homogeneity to more than one species (Simmon *et al.* 2008). In this study, the most variation was found in isolates that had a v1+v2 16S rRNA sequence most homogeneous to a sequence of a type strain belonging to the *S. mitis* - *S. sanguinis* group.

Variant 16S rRNA alleles in a single strain have been detected in many reports (Acinas *et al.* 2004; Clarridge 2004), and different 16S rRNA sequences were also detected in this study in the genomic sequences of *S. gordonii* str. Challis substr. CH1 and *S. thermophilus* LMD-9. In addition, variant 16S rRNA alleles were recently shown to be relatively common in streptococcal isolates belonging to the *S. mitis* group, as even isolates deposited to a culture collection were found to contain more than one 16S rRNA sequence, and this variation was considered as direct evidence of HGT between *S. mitis* and *S. oralis* clusters (Kilian *et al.* 2008). HGT of the 16S rRNA was also suggested in an earlier study in isolates belonging to the *S. anginosus* group (Schouls *et al.* 2003). These findings support the presence of heterogeneous sequences also among the alpha-haemolytic streptococcal isolates of this study. The results of this study further indicate that the normal microbiota isolates have acquired new genetic material more often than the disease-causing microbes. This is not surprising as the isolates residing in the normal microbiota have always other bacterial isolates around them from which new genetic material can be acquired.

The intra-isolate sequence heterogeneity in 16S rRNA sequences has also been regarded as an indication that the 16S rRNA sequence could not be used for streptococcal identification(Kilian *et al.* 2008). However, the heterogeneity detected in this study was minimal and majority of the isolates could be identified at least to the streptococcal group level despite the heterogeneous sequences. With routine Sanger sequencing, the subtle differences among the 16S alleles would probably not be

detected (Clarridge 2004), but this kind of heterogeneities are easily revealed using pyrosequencing. In addition, HGT is thought to happen among very closely related bacterial clusters, which does not affect the group-level identification of isolates, which is often considered sufficient in patient care.

The heterogeneous sequences could also be due to impure culture, and isolates carrying only few heterogeneous bases can occasionally be separated with very careful analysis (Clarridge 2004). Whether this is the case also in our isolates containing heterogeneous sequences is not known but if there were more than one bacterium growing, they must be very closely related since we tried to make pure cultures of the isolates several times, and at least on blood agar plate the growth was homogeneous. Especially the blood culture isolate that had heterogeneous sequences in both regions could however be a mixed culture. Also the possibility of a PCR-generated sequencing error can not be ruled out although a proofreading DNA polymerase was used in the template preparation PCR (Eckert and Kunkel 1991).

6.1.4 Comparison of different methods for the identification of streptococci

In this study, species-level agreement in the pyrosequencing and Vitek 2 results was found in 30/124 (24.2%) and group-level agreement was reached in 74 (60.0%) isolates. The low species-level concordance of this study mainly reflects the difficulty of identifying *S. mitis* - *S. sanguinis* group members to the species level by both methods. Only 23/69 (33.3 %) of the isolates identified to the *S. mitis* - *S. sanguinis* group by pyrosequencing were identified to the species level. By Vitek 2, species-level results can not be reached for the *S. mitis* group members since *S. mitis* group members are identified as *S. mitis / S. oralis*.

Previously, Innings *et al.* (2005) compared pyrosequencing of two regions of the *rnpB* of 113 streptococcal isolates to the commercial VITEK 2 with fluorimetric GP card and Rapid ID 32 Strep (RID32) methods. By the *rnpB* method, the identification result was mostly the species that had the highest degree of homology to the tested strain (Innings *et al.* 2005). In contrast, the degree of homology was considered in the present study and the results were considered as group-level results only if the sequence differed by more than one nucleotide from the closest type strain sequence. With the *rnpB* pyrosequencing method, analysis of 105 (92.9 %) isolates resulted in single-species identification, and all the isolates for which a clear species-level identification was not obtained were α-haemolytic (Innings *et al.* 2005).

In the study of Innings *et al.*, all the methods gave the same result at the species level for 58 (51.3 %) isolates, 85 (75.2 %) isolates had concordant results by pyrosequencing and VITEK 2 and 88 (77.8 %) with pyrosequencing and RID32. The higher level of concordance is also possibly explained by the small number of isolates (n=34) belonging to the *S. mitis* - *S. sanguinis* group. Of the isolates belonging to the *S. mitis* – *S. sanguinis* group, 14 (41.1%) were identified to the same species. Moreover, 11 of these were *S. pneumoniae* isolates that were also found to have homogeneous sequences in the v1 and v2 regions of the 16S rRNA analysed in this study.

In other studies, similar results have been obtained: when sequence analysis of the full-length *rnpB* gene was compared to API20 Strep system, 39/76 (51%) of viridans group streptococcal isolates had results concordant at the species level (Westling *et al.* 2008). When sequencing of *rpoB, tuf*, and 16S rRNA was compared to phenotypic identification using commercially available panel on a Microscan walkaway PC-21 instrument (Dade Behring, Derrfield, IL, USA), the results were concordant at the group level for 69/94 (72.6 %) VGS isolates that had caused endocarditis (Simmon *et al.* 2008). These results emphasize the fact that the identification is based on different mechanisms that are not tightly dependent on each other; the biochemical profiles of isolates may be identical although their genomic sequences are different and vice versa.

The identification results with low quality or homology strongly indicate that the databases used for the identification of alpha-haemolytic streptococci in this and other studies are insufficient if species-level identification of alpha-haemolytic streptococci is required. In addition, the pyrosequencing analysis reveals that there is considerable variation within a single VITEK result. Furthermore, the low pyrosequencing scores show that there is more variation among the 16S rRNA sequences of streptococcal isolates than in the sequences of accepted type strains. In current taxonomic thinking (Tindall *et al.* 2010), this indicates the potential presence of new streptococcal species. Despite the insufficient databases however, 16S rRNA sequencing-based methods can be used reliably to obtain at least a clue of the bacterial identity.

6.2 Detection of resistance mutations by pyrosequencing (III, IV)

Heterogeneous sequences, as the macrolide resistance mutations at the positions 2058-2059 of the 23S rRNA and the ESBL mutations of the SHV gene are difficult to determine. Accordingly, labour-intensive methods as single strand conformation polymorphism (SSCP) (M'Zali *et al.* 1998) have been used to determine the exact sequence combinations of an isolate, but these methods are not suitable for routine use. Pyrosequencing is intrinsically suitable for determining heterogeneous sequences. Pyrosequencing was first applied to the detection of mutations in the 23S rRNA causing linezolid resistance in enterococci in 2003 (Sinclair *et al.* 2003).

Heterogeneous sequences can also be determined using Sanger sequencing, but specifying the exact sequence combination of isolates containing more than one sequence requires cloning and sequencing the single molecules (al Naiemi *et al.* 2006). The individual sequences can in principle be determined also from the ambiguous electropherograms, but because the peak heights in Sanger sequencing are not quantitative, very careful analysis as well as taking into account the quality values of the sequencing results is required (al Naiemi *et al.* 2006; Schuurman *et al.* 1999). Software for analysing ambiguous electropherograms have recently been developed and have been used for analysing mixed electropherograms in the 16S rRNA sequences of samples containing more than one bacterial species (Kommedal *et al.* 2008). With such software, heterogeneous resistance mutations should be easily resolved, but the pyrosequencing procedure is much simpler and faster to perform at present. In

addition, pyrosequencing has been found to estimate more reliably the proportions of different sequence types than Sanger sequencing (Lindström *et al.* 2004).

Real-time PCR with specific probes (Randegger and Hächler 2001; Schabereiter-Gurtner *et al.* 2004) or high resolution melting curve analysis (Wolff *et al.* 2008) have been applied to detect macrolide resistance mutations in the 23S rRNA or different SHV types. However, only bacteria containing homogenous 23S rRNA or a single SHV copy were studied in the real-time PCR studies, but probably the methods could also be used for detecting an isolate containing heterogeneous target sequences. Most recently, different ESBL genes have been analysed using DNA-chips (Endimiani *et al.* 2010; Leinberger *et al.* 2010) and also *K. pneumoniae* isolates containing two SHV copies could be successfully typed with a commercial chip (Endimiani *et al.* 2010). However, the results obtained with real-time PCR or DNA chips are not as clear and reliable as pyrosequencing in which the exact sequence result can be seen. In addition, sequence information is imperative also for designing probes for real-time PCR or DNA-chips.

6.2.1 Detection of mutations causing resistance to macrolides (III)

Sinclair et al noticed that the number of mutated 23S rRNA copies could be estimated using pyrosequencing, and good correlation was found between the mutated alleles and the level of resistance (Sinclair *et al.* 2003). In addition, pyrosequencing has been applied to detection of several other mutations causing resistance to various antibiotics (Naas *et al.* 2007; Naas *et al.* 2006; Poirel *et al.* 2006). In this study, pyrosequencing was found to be able to detect and quantitate mutations that reside in successive nucleotides.

Pyrosequencing of the macrolide resistance is very useful in for example diagnosing *Helicobacter pylori* infection since macrolide therapy is the most common therapy in *H. pylori* infection and *H. pylori* isolates often have developed resistance by acquiring the mutations corresponding to the *E. coli* positions 2058-2059 of the 23S rRNA. Especially the possibility of identifying *H. pylori* and detecting its macrolide resistance directly from a clinical sample is advantageous in the diagnostics of *H. pylori* due to the fastidious growth conditions and slow growth of the organism (Moder *et al.* 2007). Also in other organisms such as mycobacteria that require long culturing before phenotypic susceptibility data is available, pyrosequencing is very useful for determining the resistance status (Randegger and Hächler 2001).

6.2.2 Typing of SHV genes by pyrosequencing (IV)

Most *Enterobacteriaceae* carry a chromosomal β-lactamase gene, which in *K. pneumoniae* is SHV. The chromosomal gene is usually wild type, non-ESBL. In addition to the chromosomal genes, enterobacterial isolates often acquire ESBL type β-lactamases that usually reside in plasmids. The presence of two β-lactamase genes is often problematic for molecular methods as the detection of ESBL type may be difficult due to the interfering wild-type signal from the chromosomal copy.

In this study, pyrosequencing of the positions 35, 238, and 240 was used to resolve the SHV copies of *K. pneumoniae* isolates. Although all SHV types cannot be detected by studying these positions, most of the ESBL SHV enzymes have a mutation at the position 238 and/or 240 (Jacoby and Bush 2010) and are identified as ESBLs by the methods developed in this study. In addition, vast majority of the SHV types found in a recent Finnish study (Nyberg *et al.* 2007) are covered.

The pyrosequencing dispensation orders of this study were designed to allow easy visual interpretation of the mutation status of the positions 35, 238 and 240 and the presence of more than one SHV copy. The dispensation orders of this study generated pyrograms in which the mutation status and quantification of the different SHV copies could be verified from several different peaks, which has been found useful also in earlier studies (Lindström *et al.* 2004). In addition, visual estimation has been found more precise than the software-based. This is probably because in software, the signals are thought to descend much more towards the end of the run than they presently do (Doostzadeh *et al.* 2008). Hopefully, this is a temporary flaw of the pyrosequencing software that will be resolved.

The use of the automatic SNP pyrosequencing format facilitates the analysis of especially the pyrograms of samples containing a mixed sequence as the software interprets the pyrograms and gives a percentage value for each sequence. The automatic analysis is very useful especially when the number of analyses is high (Jones *et al.* 2008). The SNP protocol of pyrosequencing was used for the determination and quantification of SHV type by determining the sequences of nucleotide positions 8, 35, 43, 119, 156, 179, 205, 238, and 240 using as many as eight pyrosequencing reactions and four different PCR products to obtain also the sequence opposite strand at some positions (Jones *et al.* 2008). However, the SNP software was not able to determine the percentage for the position 240 due to the two sequence options for SHV-1 at that position (Jones *et al.* 2008).

Whether using the SHV typing method of this study or the method of Jones et al, the exact SHV type cannot be determined by pyrosequencing and sequencing the entire gene seems to be the solution. The pyrosequencing methods can be used to rapidly identify SHV types spreading in a hospital, for instance. The methods are also useful in epidemiological studies to rapidly determine what kind of bacteria are spreading as the resistance phenotype does not reveal the genetic relatedness of the isolates.

Distribution of SHV types (IV)

In this study, more than one SHV copy was present in 72 of 106 (67.9%) studied Finnish *K. pneumoniae* isolates, and 65 of these were found to contain an ESBL copy by pyrosequencing also supporting the view that the ESBL SHVs usually reside on plasmids. In the study of Jones *et al*., isolates with three, potentially even four SHV copies were found (Jones *et al.* 2008) which supports our finding of an isolate carrying three SHV copies.

In this study, SHV-1 and SHV-11 were found in 42 (39.6%) and 53 (50%) Finnish *K. pneumoniae* isolates, respectively. Interestingly, SHV-11 was not found in 121 German *K. pneumoniae* isolates using a microarray method (Leinberger *et al.* 2010).

In a recent study, *K. pneumoniae* isolates with non-ESBL-SHV genes were induced to ESBL type *in vitro*, and the chromosomal copy did not mutate to an ESBL type. On the contrary, the SHVs converted easily to ESBL type in strains carrying a plasmid-borne SHV. Acquiring the ESBL type mutations under selection pressure seemed to require insertion element IS26 either 2 kb upstream of the SHV gene or in the promoter region. Consequently, presence of a non-ESBL SHV copy on a plasmid in *K. pneumoniae* confers high risk for the isolate to become an ESBL under selective pressure (Hammond *et al.* 2008). Thus, the detection of SHV-1 and SHV-11 in the same isolate contains a warning that the isolate is prone to become ESBL type. In this study, five *K. pneumoniae* isolates were found to carry SHV-1 and SHV-11 and are thus in danger of mutating to an ESBL type under selection pressure.

7 SUMMARY AND CONCLUSIONS

Identification of bacterial isolate and determining its resistance profile are the major tasks of a clinical microbiology laboratory. In this study, pyrosequencing-based methods for the identification of streptococci and for the detection of mutations causing antimicrobial resistance were developed. The methods are mainly useful in the surveillance of microbes and microbial infections in reference laboratories but in principle can be used in routine diagnostics.

The alpha-haemolytic streptococci are very difficult to identify and differentiate by traditional phenotypic methods but also by genotypic methods. Currently, the alpha-haemolytic streptococci except *S. pneumoniae* are rarely identified in routine laboratories. This is probably partly due to the tedious identification but also due to the existence of alpha-haemolytic streptococci in the human normal microbiota and insufficient knowledge on infections caused by alpha-haemolytic streptococci. In this study, great variation in streptococcal 16S rRNA sequences was found, and due to the variation, the identification of most alpha-haemolytic streptococcal isolates could be performed only to the group level. However, the method developed in this study is able to identify more streptococcal isolates than the phenotypic methods because an identification sequence result was obtained for all the studied isolates by pyrosequencing and the vast majority of the isolates could be identified at least to a streptococcal group.

The pyrosequencing methods developed for the detection of mutations causing antimicrobial resistance were found to be reliable, fast and give more information on the types of mutations when compared to traditional sequencing or probe-based methods. The method developed for the detection of mutations causing macrolide resistance at the positions 2058-2059 of the 23S rRNA was found to be quantitative; the amount of mutated 23S rRNA copies could be determined either simply visually or more specifically by analysing the pyrosequencing peak heights. The isolates having ambiguous sequence results at the amino acid positions 35 and/or 238-240 of the SHV molecule by traditional sequencing could be resolved by the pyrosequencing method developed in this study. The main advantage of pyrosequencing in resolving ambiguous sequences is the ability to determine the dispensation order specifically for each application or even for a single sample.

In conclusion, pyrosequencing was found to be a reliable, fast, and cost-effective sequencing method with which various sequencing tasks can be resolved.

8 ACKNOWLEDGEMENTS

This study was carried out at the Antimicrobial Resistance Unit, National Institute for Health and Welfare in Turku and at the Department of Medical Microbiology and Immunology, University of Turku.

I thank my supervisors docent Jari Jalava and professor Pentti Huovinen for the opportunity to do my thesis project under their supervision. Jari Jalava's enthusiasm and broad knowledge are admirable. I always got a thorough answer to my questions and usually something additional to think further. Pentti Huovinen is thanked for valuable comments and clear views during preparation of the manuscripts and especially this thesis.

Professor Matti Viljanen, former head of the Department of Medical Microbiology and Immunology, is acknowledged for accepting this thesis to be published from the Department of Medical Microbiology and Immunology, University of Turku.

Docent Risto Vuento, MD, PhD and Baback Gharizadeh, PhD are acknowledged for thorough review and constructive comments on this thesis. In addition, they are thanked for suggesting important additions to the thesis.

I'm thankful to my co-authors of the study I, especially late Helena Seppälä, MD, PhD for accepting pyrosequencing to be used in the study and for selecting me as the second author of the paper. Kaisu Rantakokko-Jalava, MD, PhD and Olli Meurman, MD, PhD are specially thanked for collaboration in the streptococcal identification project. Sofia Forssten, PhD is thanked for her significant contribution in the SHV project.

I'm grateful to Johanna Mäkinen, PhD for asking me more than ten years ago whether I would be interested in carrying out a thesis project at the National Public Health Institute. My career in THL would not have happened without you! Docent Hanna Soini and Merja Marjamäki are thanked for the opportunity to continue working at the National Institute for Health and Welfare, at the Mycobacterial laboratory. They are also appreciated for the possibility to complete this thesis work along with my main tasks.

Docent Antti Hakanen is thanked for giving practical advice in several issues and for keeping the spirit by always having an anecdote at hand. I appreciate Erkki Nieminen for help in preparation of many problematic pyrogram figures during the thesis project.

Tuula Randell and Mari Virta are thanked for excellent technical assistance. In addition, several people have assisted me "behind the scenes". Tarja Laustola and Katri Kylä-Mattila are thanked for preparation of the reagents and Maritta Kantanen and Tarja Boman for taking care of the equipment. The ladies in the office, Virpi Aaltonen, Jaana Halme, Pirjo Ketonen and Paula Kiiskilä are thanked for help with various bureaucratic issues.

I'm grateful to past and present PhD students of our group, most of whom are also my past roommates, Miika Bergman, Sofia Forssten, PhD, Marianne Gunell, PhD, Mirva Lehtopolku, Laura Lindholm, Merja Rantala, PhD, Annakaisa Suominen, Kati

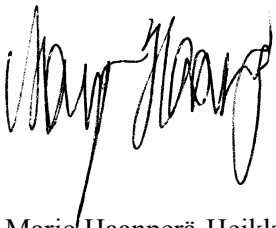
Vuorenoja, and Monica Österblad, PhD for company, support, and discussions on everything. Past and present PhD students of other groups and the whole staff of THL Mikro is thanked for a creating a good working atmosphere and for sometimes hilarious discussions in the coffee room.

My friends are thanked for kindly reminding me of life outside this project, work, and household. I'm very grateful to Anneli Vieno for taking loving care of Helmi and Leevi during the working days. We were extremely lucky when we got her as the childminder and almost as a third grandmother for our children! Ari is thanked for being a great brother and relating to this project, he is thanked for IT-support that has been available almost 24/7. My parents Sirpa and Heikki are thanked for everything, especially for having faith in me. My parents-in-law Eija-Liisa and Jussi are thanked for their encouragement and support during the years. Both my parents and parents-in-law are thanked for help in childcare that has been available even with a short notice and that has been indispensable for the finishing of this project.

Helmi and Leevi, I promise not to spend as much time sitting at the computer as I have been lately. You are most important to me. My husband Jussi-Pekka is thanked for patience, love, and understanding during the ups and downs of our journey so far. I'm especially grateful for him for taking care of more than his share of the household chores recently.

This work was financially supported by the MICMAN program of the Academy of Finland, a special grant from Turku University Central hospital, a Special Governmental (EVO) Grant of the Hospital District of Southwest Finland, Maud Kuistila memorial foundation, Paulo foundation, and Valto Takala foundation.

Turku, April 2011



Marjo Haanperä-Heikkinen

REFERENCES

Abraham, E. P. and E. Chain (1940). An Enzyme from Bacteria able to Destroy Penicillin. *Nature* **146**: 837.

Achtman, M. and M. Wagner (2008). Microbial diversity and the genetic nature of microbial species. *Nat Rev Micro* **6**(6): 431-440.

Acinas, S. G., L. A. Marcelino, V. Klepac-Ceraj and M. F. Polz (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J Bacteriol* **186**(9): 2629-2635.

Ahmadian, A., M. Ehn and S. Hober (2006). Pyrosequencing: History, biochemistry and future. *Clin Chim Acta* **363**(1-2): 83-94.

Ahmadian, A., B. Gharizadeh, D. O'Meara, J. Odeberg and J. Lundeberg (2001). Genotyping by apyrase-mediated allele-specific extension. *Nucleic Acids Res* **29**(24): E121.

Ahmadian, A., J. Lundeberg, P. Nyrén, M. Uhlen and M. Ronaghi (2000). Analysis of the p53 tumor suppressor gene by pyrosequencing. *Biotechniques* **28**(1): 140-144, 146-147.

al Naiemi, N., K. Schipper, B. Duim and A. Bart (2006). Application of Minimal Sequence Quality Values Prevents Misidentification of the blaSHV Type in Single Bacterial Isolates Carrying Different SHV Extended-Spectrum β-Lactamase Genes. *J Clin Microbiol* **44**(5): 1896-1898.

Alvarez-Elcoro, S. and M. J. Enzler (1999). The macrolides: erythromycin, clarithromycin, and azithromycin. *Mayo Clin Proc* **74**(6): 613-634.

Ambler, R. P., A. F. Coulson, J. M. Frere, J. M. Ghuysen, B. Joris, M. Forsman, R. C. Levesque, G. Tiraby and S. G. Waley (1991). A standard numbering scheme for the class A beta-lactamases. *Biochem J* **276 (Pt 1)**: 269-270.

Andrewes, F. W. and T. J. Horder (1906). A study of the streptococci pathogenic for man. *Lancet* **2**: 708–713.

Anhalt, J. P. and C. Fenselau (1975). Identification of bacteria using mass spectrometry. *Anal Chem* **47**(2): 219-225.

Arbique, J. C., C. Poyart, P. Trieu-Cuot, G. Quesne, M. d. G. S. Carvalho, A. G. Steigerwalt, R. E. Morey, D. Jackson, R. J. Davidson and R. R. Facklam (2004). Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. nov. *J Clin Microbiol* **42**(10): 4686-4696.

Ayers, S. H. and C. S. Mudge (1922). The streptococci of the bovine udder. IV. Studies of the streptococci. *J Infect Dis* **31**: 40.

Baba, Y., K. Nosho, K. Shima, C. Huttenhower, N. Tanaka, A. Hazra, E. L. Giovannucci, C. S. Fuchs and S. Ogino (2010). Hypomethylation of the IGF2 DMR in Colorectal Tumors, Detected by Bisulfite Pyrosequencing, is Associated with Poor Prognosis. *Gastroenterology* **139**(6): 1855-1864.

Babini, G. S. and D. M. Livermore (2000). Are SHV β-Lactamases Universal in *Klebsiella pneumoniae*? *Antimicrob Agents Chemother* **44**(8): 2230.

Balm, M. N. D., H. T. Truong, A. S. Choudhary, G. M. Robinson and T. K. Blackmore (2006). *Streptococcus gallinaceus* bacteraemia in an abattoir worker presenting with a febrile illness. *J Med Microbiol* **55**(7): 957-959.

Baltimore, R. S. (2007). Consequences of Prophylaxis for Group B Streptococcal Infections of the Neonate. *Semin Perinatol* **31**(1): 33-38.

Barnes, E. M., C. S. Impey, B. J. Stevens and J. L. Peel (1977). *Streptococcus pleomorphus* sp.nov.: an anaerobic streptococcus isolated mainly from the caeca of birds. *J Gen Microbiol* **102**(1): 45-53.

Bartkus, J. M., B. A. Juni, K. Ehresmann, C. A. Miller, G. N. Sanden, P. K. Cassiday, M. Saubolle, B. Lee, J. Long, A. R. Harrison, Jr. and J. M. Besser (2003). Identification of a mutation associated with erythromycin resistance in *Bordetella pertussis*: implications for surveillance of antimicrobial resistance. *J Clin Microbiol* **41**(3): 1167-1172.

Beighton, D., H. Hayday, R. R. B. Russell and R. A. Whiley (1984). *Streptococcus macacae* sp. nov.

from the dental plaque of monkeys (*Macaca fascicularis*). *Int J Syst Bacteriol* **34**: 332-335.

Bekal, S., C. Gaudreau, R. A. Laurence, E. Simoneau and L. Raynal (2006). *Streptococcus pseudoporcinus* sp. nov., a Novel Species Isolated from the Genitourinary Tract of Women. *J Clin Microbiol* **44**(7): 2584-2586.

Bergman, M., S. Huikko, P. Huovinen, P. Paakkari, H. Seppälä and R. Finnish Study Group for Antimicrobial (2006). Macrolide and Azithromycin Use Are Linked to Increased Macrolide Resistance in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* **50**(11): 3646-3650.

Bishop, C. J., D. M. Aanensen, G. E. Jordan, M. Kilian, W. P. Hanage and B. G. Spratt (2009). Assigning strains to bacterial species via the internet. *BMC Biol* **7**: 3.

Bizzini, A., K. Jaton, D. Romo, J. Bille, G. Prod'hom and G. Greub (2011). Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry as an Alternative to 16S rRNA Gene Sequencing for Identification of Difficult-To-Identify Bacterial Strains. *J Clin Microbiol* **49**(2): 693-696.

Bonnal, R. J. P., M. Severgnini, A. Castaldi, R. Bordoni, M. Iacono, A. Trimarco, A. Torella, G. Piluso, S. Aurino, G. Condorelli, G. De Bellis and V. Nigro (2010). Reliable resequencing of the human dystrophin locus by universal long polymerase chain reaction and massive pyrosequencing. *Anal Biochem* **406**(2): 176-184.

Bradford, P. A. (2001). Extended-spectrum β-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. *Clin Microbiol Rev* **14**(4): 933-951.

Buckley, M. and R. J. Roberts. (2007). Reconciling microbial systematics and genomics. The American Academy of Microbiology Reports. Accessed February 25, 2011, from http://academy.asm.org/images/stories/documents/reconcilingmicrobialsystematicsandgenomicsfull.pdf.

Burgers, P. M. and F. Eckstein (1979). A study of the mechanism of DNA polymerase I from *Escherichia coli* with diastereomeric phosphorothioate analogs of deoxyadenosine triphosphate. *J Biol Chem* **254**(15): 6889-6893.

Buriankova, K., F. Doucet-Populaire, O. Dorson, A. Gondran, J.-C. Ghnassia, J. Weiser and J.-L. Pernodet (2004). Molecular Basis of Intrinsic Macrolide Resistance in the *Mycobacterium tuberculosis* Complex. *Antimicrob Agents Chemother* **48**(1): 143-150.

Bush, K. (2001). New beta-lactamases in gram-negative bacteria: diversity and impact on the selection of antimicrobial therapy. *Clin Infect Dis* **32**(7): 1085-1089.

Canu, A., A. Abbas, B. Malbruny, F. Sichel and R. Leclercq (2004). Denaturing high-performance liquid chromatography detection of ribosomal mutations conferring macrolide resistance in gram-positive cocci. *Antimicrob Agents Chemother* **48**(1): 297-304.

Canu, A., B. Malbruny, M. Coquemont, T. A. Davies, P. C. Appelbaum and R. Leclercq (2002). Diversity of Ribosomal Mutations Conferring Resistance to Macrolides, Clindamycin, Streptogramin, and Telithromycin in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* **46**(1): 125-131.

Carvalho, M. d. G. S., M. L. Tondella, K. McCaustland, L. Weidlich, L. McGee, L. W. Mayer, A. Steigerwalt, M. Whaley, R. R. Facklam, B. Fields, G. Carlone, E. W. Ades, R. Dagan and J. S. Sampson (2007). Evaluation and Improvement of Real-Time PCR Detection Assays to *lytA*, *ply*, and *psaA* Genes for Detection of Pneumococcal DNA. *J Clin Microbiol* **45**(8): 2460-2466.

Caufield, P. W., A. P. Dasanayake, Y. Li, Y. Pan, J. Hsu and J. M. Hardin (2000). Natural History of *Streptococcus sanguinis* in the Oral Cavity of Infants: Evidence for a Discrete Window of Infectivity. *Infect Immun* **68**(7): 4018-4023.

Chang, J. Y., D. A. Antonopoulos, A. Kalra, A. Tonelli, W. T. Khalife, T. M. Schmidt and V. B. Young (2008). Decreased diversity of the fecal Microbiome in recurrent *Clostridium difficile*-associated diarrhea. *J Infect Dis* **197**(3): 435-438.

Chang, W., T. Huang and L. Chien (2002). Identification of Viridans Streptococcal Species Causing Bacterial Meningitis in Adults in Taiwan. *Eur J Clin Microbiol Infect Dis* **21**(5): 393-396.

Chaves, J., M. G. Ladona, C. Segura, A. Coira, R. Reig and C. Ampurdanes (2001). SHV-1 β-lactamase is mainly a chromosomally encoded species-specific enzyme in *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* **45**(10): 2856-2861.

Chen, C. C., L. J. Teng and T. C. Chang (2004). Identification of Clinically Relevant Viridans Group Streptococci by Sequence Analysis of the 16S-23S Ribosomal DNA Spacer Region. *J Clin Microbiol* **42**(6): 2651-2657.

Chun, J., J. H. Lee, Y. Jung, M. Kim, S. Kim, B. K. Kim and Y. W. Lim (2007). EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int J Syst Evol Microbiol* **57**(Pt 10): 2259-2261.

Clancy, J., J. Petitpas, F. Dib-Hajj, W. Yuan, M. Cronan, A. V. Kamath, J. Bergeron and J. A. Retsema (1996). Molecular cloning and functional analysis of a novel macrolide-resistance determinant, *mefA*, from *Streptococcus pyogenes*. *Mol Microbiol* **22**(5): 867-879.

Clarridge, J. E., III (2004). Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clin Microbiol Rev* **17**(4): 840-862.

Collins, M., R. Hutson, E. Falsen, N. Nikolaitchouk, L. LaClaire and R. Facklam (2000). An unusual *Streptococcus* from human urine, *Streptococcus urinalis* sp. nov. *Int J Syst Evol Microbiol* **50**(3): 1173-1178.

Collins, M., R. Hutson, L. Hoyles, E. Falsen, N. Nikolaitchouk and G. Foster (2001). *Streptococcus ovis* sp. nov., isolated from sheep. *Int J Syst Evol Microbiol* **51**(3): 1147-1150.

Collins, M. D., R. A. Hutson, E. Falsen, E. Inganas and M. Bisgaard (2002). *Streptococcus gallinaceus* sp. nov., from chickens. *Int J Syst Evol Microbiol* **52**(4): 1161-1164.

Collins, M. D., P. A. Lawson, A. Willems, J. J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe and J. A. Farrow (1994). The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol* **44**(4): 812-826.

Collins, M. D., T. Lundstrom, C. Welinder-Olsson, I. Hansson, O. Wattle, R. A. Hutson and E. Falsen (2004). *Streptococcus devriesei* sp. nov., from Equine Teeth. *Syst Appl Microbiol* **27**(2): 146-150.

Coque, T. M., F. Baquero and R. Canton (2008). Increasing prevalence of ESBL-producing *Enterobacteriaceae* in Europe. *Euro Surveill* **13**(47).

Corredoira, J. C., M. P. Alonso, J. F. García, E. Casariego, A. Coira, A. Rodriguez, J. Pita, C. Louzao, B. Pombo, M. J. López and J. Varela (2005). Clinical characteristics and significance of *Streptococcus salivarius* bacteremia and *Streptococcus bovis* bacteremia: a prospective 16-year study. *Eur J Clin Microbiol Infect Dis* **24**(4): 250-255.

Coykendall, A. L. (1989). Classification and identification of the viridans streptococci. *Clin Microbiol Rev* **2**(3): 315-328.

Daley, P., D. L. Church, D. B. Gregson and S. Elsayed (2005). Species-level molecular identification of invasive "*Streptococcus milleri*" group clinical isolates by nucleic acid sequencing in a centralized regional microbiology laboratory. *J Clin Microbiol* **43**(6): 2987-2988.

Del Grosso, M., F. Iannelli, C. Messina, M. Santagati, N. Petrosillo, S. Stefani, G. Pozzi and A. Pantosti (2002). Macrolide Efflux Genes *mef(A)* and *mef(E)* Are Carried by Different Genetic Elements in *Streptococcus pneumoniae*. *J Clin Microbiol* **40**(3): 774-778.

Delcour, A. H. (2009). Outer membrane permeability and antibiotic resistance. *Biochim Biophys Acta* **1794**(5): 808-816.

Delorme, C. (2008). Safety assessment of dairy microorganisms: *Streptococcus thermophilus*. *Int J Food Microbiol* **126**(3): 274-277.

Delorme, C., C. Poyart, S. D. Ehrlich and P. Renault (2007). Extent of Horizontal Gene Transfer in Evolution of Streptococci of the Salivarius Group. *J. Bacteriol.* **189**(4): 1330-1341.

Devriese, L., B. Pot, P. Vandamme, K. Kersters, M. Collins, N. Alvarez, F. Haesebrouck and J. Hommez (1997). *Streptococcus hyovaginalis* sp. nov. and *Streptococcus thoraltensis* sp. nov., from the genital tract of sows. *Int J Syst Bacteriol* **47**(4): 1073-1077.

Devriese, L. A., R. Kilpper-BaLz and K. H. Schleifer (1988). NOTES: *Streptococcus hyointestinalis* sp. nov. from the Gut of Swine. *Int J Syst Bacteriol* **38**(4): 440-441.

Devriese, L. A., P. Vandamme, M. D. Collins, N. Alvarez, B. Pot, J. Hommez, P. Butaye and F. Haesebrouck (1999). *Streptococcus pluranimalium* sp. nov., from cattle and other animals. *Int J Syst Bacteriol* **49 Pt 3**: 1221-1226.

Diernhofer, K. (1932). Aesculinbouillion als Hilfsmittel für die Differenzierung von Euter und Milchstreptokokken bei Massenuntersuchungen. *Milchwirtschaftl. Forsch.* **13**: 368-378.

Doern, C. D. and C.-A. D. Burnham (2010). It's Not Easy Being Green: the Viridans Group Streptococci, with a Focus on Pediatric Clinical Manifestations. *J Clin Microbiol* **48**(11): 3829-3835.

Doostzadeh, J., S. Shokralla, F. Absalan, R. Jalili, S. Mohandessi, J. W. Langston, R. W. Davis, M. Ronaghi and B. Gharizadeh (2008). High Throughput Automated Allele Frequency Estimation by Pyrosequencing. *PLoS ONE* **3**(7): e2693.

Drancourt, M., V. Roux, P.-E. Fournier and D. Raoult (2004). *rpoB* gene sequence-based identification of aerobic gram-positive cocci of the genera *Streptococcus, Enterococcus, Gemella, Abiotrophia,* and *Granulicatella*. *J Clin Microbiol* **42**(2): 497-504.

Duarte, R. S., R. R. Barros, R. R. Facklam and L. M. Teixeira (2005). Phenotypic and genotypic characteristics of *Streptococcus porcinus* isolated from human sources. *J Clin Microbiol* **43**(9): 4592-4601.

Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson and D. A. Relman (2005). Diversity of the human intestinal microbial flora. *Science* **308**(5728): 1635-1638.

Eckert, K. A. and T. A. Kunkel (1991). DNA polymerase fidelity and the polymerase chain reaction. *Genome Res* **1**(1): 17-24.

Endimiani, A., A. M. Hujer, K. M. Hujer, J. A. Gatta, A. C. Schriver, M. R. Jacobs, L. B. Rice and R. A. Bonomo (2010). Evaluation of a Commercial Microarray System for Detection of SHV-, TEM-, CTX-M-, and KPC-Type β-Lactamase Genes in Gram-Negative Isolates. *J Clin Microbiol* **48**(7): 2618-2622.

Engberg, J., F. M. Aarestrup, D. E. Taylor, P. Gerner-Smidt and I. Nachamkin (2001). Quinolone and macrolide resistance in *Campylobacter jejuni* and *C. coli*: resistance mechanisms and trends in human isolates. *Emerg Infect Dis* **7**(1): 24-34.

England, R. and M. Pettersson (2005). Pyro Q-CpG™: quantitative analysis of methylation in multiple CpG sites by Pyrosequencing®. *Nat Meth* **2**(10).

Eriksson, J., B. Gharizadeh, T. Nordstrom and P. Nyrén (2004). Pyrosequencing trade mark technology at elevated temperature. *Electrophoresis* **25**(1): 20-27.

Euzéby, J. P. (2011, February 23). List of Prokaryotic names with Standing in Nomenclature. Accessed February 28, 2011, from http://www.bacterio.cict.fr/.

Ezaki, T., Y. Hashimoto, N. Takeuchi, H. Yamamoto, S. L. Liu, H. Miura, K. Matsui and E. Yabuuchi (1988). Simple genetic method to identify viridans group streptococci by colorimetric dot hybridization and fluorometric hybridization in microdilution wells. *J Clin Microbiol* **26**(9): 1708-1713.

Facklam, R. (2002). What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clin Microbiol Rev* **15**(4): 613-630.

Fakhrai-Rad, H., N. Pourmand and M. Ronaghi (2002). Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. *Hum Mutat* **19**(5): 479-485.

Farrow, J. A. E., J. Kruze, B. A. Phillips, A. J. Bramley and M. D. Collins (1984). Taxonomic studies on *Streptococcus bovis* and *Streptococcus equinus*: description of *Streptococcus alactolyticus* sp. nov. and *Streptococcus saccharolyticus* sp. nov. *Syst Appl Microbiol* **5**: 476-482.

Fernandez, E., V. Blume, P. Garrido, M. D. Collins, A. Mateos, L. Dominguez and J. F. Fernandez-Garayzabal (2004). *Streptococcus equi*

subsp. *ruminatorum* subsp. nov., isolated from mastitis in small ruminants. *Int J Syst Evol Microbiol* **54**(6): 2291-2296.

Finkelstein, Y., N. Marcus, R. Mosseri, Z. Bar-Sever and B. Z. Garty (2003). *Streptococcus acidominimus* infection in a child causing Gradenigo syndrome. *Int J Pediatr Otorhinolaryngol* **67**(7): 815-817.

Frank, D. N., A. L. St Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz and N. R. Pace (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* **104**(34): 13780-13785.

Friedrichs, C., A. C. Rodloff, G. S. Chhatwal, W. Schellenberger and K. Eschrich (2007). Rapid identification of viridans streptococci by mass-spectrometric discrimination. *J Clin Microbiol*: JCM.00556-00507.

Garcia, C. A., A. Ahmadian, B. Gharizadeh, J. Lundeberg, M. Ronaghi and P. Nyrén (2000). Mutation detection by pyrosequencing: sequencing of exons 5-8 of the p53 tumor suppressor gene. *Gene* **253**(2): 249-257.

Garnier, F., G. Gerbaud, P. Courvalin and M. Galimand (1997). Identification of clinically relevant viridans group streptococci to the species level by PCR. *J Clin Microbiol* **35**(9): 2337-2341.

Garrity, G. M., T. G. Lilburn, J. R. Cole, S. H. Harrison, J. Enzeby and B. J. Tindall. (2007). Taxonomic Outline of Bacteria and Archaea (TOBA). Release 7.7, 6 March 2007. . Accessed 28th February, 2011.

Gauduchon, V., L. Chalabreysse, J. Etienne, M. Celard, Y. Benito, H. Lepidi, F. Thivolet-Bejui and F. Vandenesch (2003). Molecular Diagnosis of Infective Endocarditis by PCR Amplification and Direct Sequencing of DNA from Valve Tissue. *J Clin Microbiol* **41**(2): 763-766.

Gharizadeh, B., M. Akhras, N. Nourizad, M. Ghaderi, K. Yasuda, P. Nyrén and N. Pourmand (2006). Methodological improvements of pyrosequencing technology. *J Biotechnol* **124**(3): 504-511.

Gharizadeh, B., J. Eriksson, N. Nourizad, T. Nordstrom and P. Nyrén (2004). Improvements in Pyrosequencing technology by employing Sequenase polymerase. *Anal Biochem* **330**(2): 272-280.

Gharizadeh, B., M. Kalantari, C. A. Garcia, B. Johansson and P. Nyrén (2001). Typing of human papillomavirus by pyrosequencing. *Lab Invest* **81**(5): 673-679.

Gharizadeh, B., T. Nordstrom, A. Ahmadian, M. Ronaghi and P. Nyrén (2002). Long-read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer. *Anal Biochem* **301**(1): 82-90.

Gharizadeh, B., A. Ohlin, P. Molling, A. Backman, B. Amini, P. Olcen and P. Nyrén (2003). Multiple group-specific sequencing primers for reliable and rapid DNA sequencing. *Mol Cell Probes* **17**(4): 203-210.

Glazunova, O. O., D. Raoult and V. Roux (2006). *Streptococcus massiliensis* sp. nov., isolated from a patient blood culture. *Int J Syst Evol Microbiol* **56**(5): 1127-1131.

Gniadkowski, M. (2008). Evolution of extended-spectrum beta-lactamases by mutation. *Clin Microbiol Infect* **14 Suppl 1**: 11-32.

Gould, I. M. (2010). Coping with antibiotic resistance: the impending crisis. *Int J Antimicrob Agents* **36 Suppl 3**: S1-2.

Grahn, N., M. Olofsson, K. Ellnebo-Svedlund, H. J. Monstein and J. Jonasson (2003). Identification of mixed bacterial DNA contamination in broad-range PCR amplification of 16S rDNA V1 and V3 variable regions by pyrosequencing of cloned amplicons. *FEMS Microbiol Lett* **219**(1): 87-91.

Gruber, J. D., P. B. Colligan and J. K. Wolford (2002). Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing. *Hum Genet* **110**(5): 395-401.

Hæggman, S., S. Löfdahl and L. G. Burman (1997). An allelic variant of the chromosomal gene for class A beta-lactamase K2, specific for *Klebsiella pneumoniae*, is the ancestor of SHV-1. *Antimicrob Agents Chemother* **41**(12): 2705-2709.

Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold and R. Knight (2008). Error-correcting barcoded

primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**(3): 235-237.

Hammond, D. S., T. Harris, J. Bell, J. Turnidge and P. M. Giffard (2008). Selection of SHV Extended-Spectrum-β-Lactamase-Dependent Cefotaxime and Ceftazidime Resistance in *Klebsiella pneumoniae* Requires a Plasmid-Borne bla_{SHV} Gene. *Antimicrob Agents Chemother* **52**(2): 441-445.

Hanage, W. P., C. Fraser and B. G. Spratt (2006). Sequences, sequence clusters and bacterial species. *Philos Trans R Soc Lond B Biol Sci* **361**(1475): 1917-1927.

Herrero, I. A., M. S. Rouse, K. E. Piper, S. A. Alyaseen, J. M. Steckelberg and R. Patel (2002). Reevaluation of *Streptococcus bovis* endocarditis cases from 1975 to 1985 by 16S ribosomal DNA sequence analysis. *J Clin Microbiol* **40**(10): 3848-3850.

Hoffmann, C., N. Minkah, J. Leipzig, G. Wang, M. Q. Arens, P. Tebas and F. D. Bushman (2007). DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucl Acids Res* **35**(13): e91.

Hoshino, T., T. Fujiwara and M. Kilian (2005). Use of phylogenetic and phenotypic analyses to identify nonhemolytic streptococci isolated from bacteremic patients. *J Clin Microbiol* **43**(12): 6073-6085.

Hoshino, T., M. Kawaguchi, N. Shimizu, N. Hoshino, T. Ooshima and T. Fujiwara (2004). PCR detection and identification of oral streptococci in saliva samples using GTF genes. *Diagn Microbiol Infect Dis* **48**(3): 195-199.

Hugenholtz, P. and T. Huber (2003). Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *Int J Syst Evol Microbiol* **53**(1): 289-293.

Håvarstein, L. S., R. Hakenbeck and P. Gaustad (1997). Natural competence in the genus *Streptococcus*: evidence that streptococci can change pherotype by interspecies recombinational exchanges. *J. Bacteriol.* **179**(21): 6589-6594.

Innings, Å., M. Krabbe, M. Ullberg and B. Herrmann (2005). Identification of 43 *Streptococcus* species by pyrosequencing analysis of the *rnpB* gene. *J Clin Microbiol* **43**(12): 5983-5991.

Jacobs, J., C. Schot and L. Schouls (2000). The *Streptococcus anginosus* species comprises five 16S rRNA ribogroups with different phenotypic characteristics and clinical relevance. *Int J Syst Evol Microbiol* **50**(3): 1073-1079.

Jacoby, G. A. and L. S. Munoz-Price (2005). The New Beta-Lactamases. *N Engl J Med* **352**(4): 380-391.

Jacoby, J. A. and K. Bush. (2010, February 17). Amino Acid Sequences for TEM, SHV and OXA Extended-Spectrum and Inhibitor Resistant β-Lactamases. Accessed February 24, 2010, from http://www.lahey.org/Studies/.

Jalava, J., M. Vaara and P. Huovinen (2004). Mutation at the position 2058 of the 23S rRNA as a cause of macrolide resistance in *Streptococcus pyogenes*. *Ann Clin Microbiol Antimicrob* **3**(1): 5.

Jedrzejas, M. J. (2001). Pneumococcal Virulence Factors: Structure and Function. *Microbiol Mol Biol Rev* **65**(2): 187-207.

Jonasson, J., M. Olofsson and H.-J. Monstein (2002). Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16S rDNA fragments. *APMIS* **110**(3): 263-272.

Jones, C. H., A. Ruzin, M. Tuckman, M. A. Visalli, P. J. Petersen and P. A. Bradford (2008). Pyrosequencing Using the Single Nucleotide Polymorphism Protocol for Rapid determination of TEM and SHV-type Extended Spectrum β-lactamases in Clinical Isolates and Identification of the Novel β-lactamase genes blaSHV-48, blaSHV-105, and blaTEM-155. *Antimicrob Agents Chemother* **53**(3): 977-986.

Jordan, J. A., A. R. Butchko and M. Beth Durso (2005). Use of Pyrosequencing of 16S rRNA Fragments to Differentiate between Bacteria Responsible for Neonatal Sepsis. *J Mol Diagn* **7**(1): 105-110.

Joseph, S. J. and T. D. Read (2010). Bacterial population genomics and infectious disease diagnostics. *Trends Biotechnol* **28**(12): 611-618.

Karamohamed, S., J. Nilsson, K. Nourizad, M. Ronaghi, B. Pettersson and P. Nyrén (1999). Production, purification, and luminometric analysis of recombinant *Saccharomyces cerevisiae* MET3 adenosine triphosphate sulfurylase expressed in *Escherichia coli*. *Protein Expr Purif* **15**(3): 381-388.

Kawamura, Y., X. Hou, Y. Todome, F. Sultana, K. Hirose, S. Shu, T. Ezaki and H. Ohkuni (1998). *Streptococcus peroris* sp. nov. and *Streptococcus infantis* sp. nov., new members of the *Streptococcus mitis* group, isolated from human clinical specimens. *Int J Syst Bacteriol* **48**(3): 921-927.

Kawamura, Y., X. G. Hou, F. Sultana, H. Miura and T. Ezaki (1995). Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *Int J Syst Bacteriol* **45**(2): 406-408.

Kawamura, Y., Y. Itoh, N. Mishima, K. Ohkusu, H. Kasai and T. Ezaki (2005). High genetic similarity of *Streptococcus agalactiae* and *Streptococcus difficilis*: *S. difficilis* Eldar et al. 1995 is a later synonym of *S. agalactiae* Lehmann and Neumann 1896 (Approved Lists 1980). *Int J Syst Evol Microbiol* **55**(2): 961-965.

Kawamura, Y., R. A. Whiley, S.-E. Shu, T. Ezaki and J. M. Hardie (1999). Genetic approaches to the identification of the mitis group within the genus *Streptococcus*. *Microbiology* **145**(9): 2605-2613.

Kearns, A. M., J. Wheeler, R. Freeman, P. R. Seiders, J. Perry, A. M. Whatmore and C. G. Dowson (2000). Pneumolysin Detection Identifies Atypical Isolates of *Streptococcus pneumoniae*. *J Clin Microbiol* **38**(3): 1309-1310.

Kellogg, J. A., D. A. Bankert, C. J. Elder, J. L. Gibbs and M. C. Smith (2001). Identification of *Streptococcus pneumoniae* Revisited. *J Clin Microbiol* **39**(9): 3373-3375.

Kilian, M., L. Mikkelsen and J. Henrichsen (1989a). Replacement of the Type Strain of *Streptococcus mitis*: Request for an Opinion. *Int J Syst Bacteriol* **39**(4): 498-499.

Kilian, M., L. Mikkelsen and J. Henrichsen (1989b). Taxonomic Study of Viridans Streptococci: Description of *Streptococcus gordonii* sp. nov. and Emended Descriptions of *Streptococcus sanguis* (White and Niven 1946), *Streptococcus oralis* (Bridge and Sneath 1982), and *Streptococcus mitis* (Andrewes and Horder 1906). *Int J Syst Bacteriol* **39**(4): 471-484.

Kilian, M., K. Poulsen, T. Blomqvist, L. S. Havarstein, M. Bek-Thomsen, H. Tettelin and U. B. Sorensen (2008). Evolution of *Streptococcus pneumoniae* and its close commensal relatives. *PLoS ONE* **3**(7): e2683.

Kirschner, P. and E. C. Bottger (1998). Species identification of mycobacteria using rDNA sequencing. *Methods Mol Biol* **101**: 349-361.

Kliebe, C., B. A. Nies, J. F. Meyer, R. M. Tolxdorff-Neutzling and B. Wiedemann (1985). Evolution of plasmid-coded resistance to broad-spectrum cephalosporins. *Antimicrob Agents Chemother* **28**(2): 302-307.

Kommedal, Ø., B. Karlsen and Ø. Sæbø (2008). Analysis of mixed sequencing chromatograms and its application in direct 16S rDNA sequencing of poly-microbial samples. *J Clin Microbiol* **46**: 3766-3771.

Kruckeberg, K. E. and S. N. Thibodeau (2004). Pyrosequencing Technology as a Method for the Diagnosis of Multiple Endocrine Neoplasia Type 2. *Clin Chem* **50**(3): 522-529.

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

Köhler, W. (2007). The present state of species within the genera *Streptococcus* and *Enterococcus*. *Int J Med Microbiol* **297**(3): 133-150.

Lancefield, R. C. (1933). A serological differentiation of human and other groups of hemolytic streptococci *J Exp Med* **57**(4): 571-595.

Langaee, T. and M. Ronaghi (2005). Genetic variation analyses by Pyrosequencing. *Mutat Res* **573**(1-2): 96-102.

Lascols, C., D. Lamarque, J.-M. Costa, C. Copie-Bergman, J.-M. Le Glaunec, L. Deforges, C.-J.

Soussy, J.-C. Petit, J.-C. Delchier and J. Tankovic (2003). Fast and Accurate Quantitative Detection of *Helicobacter pylori* and Identification of Clarithromycin Resistance Mutations in *H. pylori* Isolates from Gastric Biopsy Specimens by Real-Time PCR. *J Clin Microbiol* **41**(10): 4573-4577.

Lawson, P. A., G. Foster, E. Falsen and M. D. Collins (2005a). *Streptococcus marimammalium* sp. nov., isolated from seals. *Int J Syst Evol Microbiol* **55**(Pt 1): 271-274.

Lawson, P. A., G. Foster, E. Falsen, N. Davison and M. D. Collins (2004). *Streptococcus halichoeri* sp. nov., isolated from grey seals (*Halichoerus grypus*). *Int J Syst Evol Microbiol* **54**(5): 1753-1756.

Lawson, P. A., G. Foster, E. Falsen, S. J. Markopoulos and M. D. Collins (2005b). *Streptococcus castoreus* sp. nov., isolated from a beaver (Castor fiber). *Int J Syst Evol Microbiol* **55**(Pt 2): 843-846.

Lee, Y. H., B. Cho, I. K. Bae, C. L. Chang and S. H. Jeong (2006). *Klebsiella pneumoniae* strains carrying the chromosomal SHV-11 beta-lactamase gene produce the plasmid-mediated SHV-12 extended-spectrum beta-lactamase more frequently than those carrying the chromosomal SHV-1 beta-lactamase gene. *J Antimicrob Chemother* **57**(6): 1259-1261.

Leinberger, D. M., V. Grimm, M. Rubtsova, J. Weile, K. Schroppel, T. A. Wichelhaus, C. Knabbe, R. D. Schmid and T. T. Bachmann (2010). Integrated Detection of Extended-Spectrum-Beta-Lactam Resistance by DNA Microarray-Based Genotyping of TEM, SHV, and CTX-M Genes. *J Clin Microbiol* **48**(2): 460-471.

Lindholm, L. and H. Sarkkinen (2004). Direct Identification of Gram-Positive Cocci from Routine Blood Cultures by Using AccuProbe Tests. *J Clin Microbiol* **42**(12): 5609-5613.

Lindström, A., J. Odeberg and J. Albert (2004). Pyrosequencing for Detection of Lamivudine-Resistant Hepatitis B Virus. *J Clin Microbiol* **42**(10): 4788-4795.

Long, S. S. and R. M. Swenson (1976). Determinants of the developing oral flora in normal newborns. *Appl Environ Microbiol* **32**(4): 494-497.

Lun, Z. R., Q. P. Wang, X. G. Chen, A. X. Li and X. Q. Zhu (2007). *Streptococcus suis*: an emerging zoonotic pathogen. *Lancet Infect Dis* **7**(3): 201-209.

Lundström, T. S., G. G. Dahlén and O. S. Wattle (2007). Caries in the infundibulum of the second upper premolar tooth in the horse. *Acta Vet Scand* **49**(1): 10.

M'Zali, F. H., J. Heritage, D. M. Gascoyne-Binzi, A. M. Snelling and P. M. Hawkey (1998). PCR single strand conformational polymorphism can be used to detect the gene encoding SHV-7 extended-spectrum beta-lactamase and to identify different SHV genes within the same strain. *J Antimicrob Chemother* **41**(1): 123-125.

Mahlen, S. D. and J. E. Clarridge, III (2009). Thumb Infection Caused by *Streptococcus pseudoporcinus*. *J Clin Microbiol* **47**(9): 3041-3042.

Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt and J. M. Tiedje (2001). The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* **29**(1): 173-174.

Manachini, P. L., S. H. Flint, L. J. Ward, W. Kelly, M. G. Fortina, C. Parini and D. Mora (2002). Comparison between *Streptococcus macedonicus* and *Streptococcus waius* strains and reclassification of *Streptococcus waius* (Flint et at. 1999) as *Streptococcus macedonicus* (Tsakalidou et al. 1998). *Int J Syst Evol Microbiol* **52**(Pt 3): 945-951.

Mankin, A. S. (2008). Macrolide myths. *Curr Opin Microbiol* **11**(5): 414-421.

Margeridon-Thermet, S., N. S. Shulman, A. Ahmed, R. Shahriar, T. Liu, C. Wang, S. P. Holmes, F. Babrzadeh, B. Gharizadeh, B. Hanczaruk, B. B. Simen, M. Egholm and R. W. Shafer (2009). Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. *J Infect Dis* **199**(9): 1275-1285.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W.

He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley and J. M. Rothberg (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**(7057): 376-380.

Marri, P. R., W. Hao and G. B. Golding (2006). Gene Gain and Gene Loss in *Streptococcus*: Is It Driven by Habitat? *Mol Biol Evol* **23**(12): 2379-2391.

Martinez-Martinez, L. (2008). Extended-spectrum beta-lactamases and the permeability barrier. *Clin Microbiol Infect* **14 Suppl 1**: 82-89.

Mashayekhi, F. and M. Ronaghi (2007). Analysis of read length limiting factors in Pyrosequencing chemistry. *Anal Biochem* **363**(2): 275-287.

Maxam, A. M. and W. Gilbert (1977). A New Method for Sequencing DNA. *PNAS* **74**(2): 560-564.

Melamede, R. J. (1989). Automatable process for sequencing nucleotide. U.S. patent 4863849

Milinovich, G. J., P. C. Burrell, C. C. Pollitt, A. Bouvet and D. J. Trott (2008). *Streptococcus henryi* sp. nov. and *Streptococcus caballi* sp. nov., isolated from the hindgut of horses with oligofructose-induced laminitis. *Int J Syst Evol Microbiol* **58**(1): 262-266.

Moder, K.-A., F. Layer, W. Konig and B. Konig (2007). Rapid screening of clarithromycin resistance in *Helicobacter pylori* by pyrosequencing. *J Med Microbiol* **56**(10): 1370-1376.

Monger, W., I. Adams, R. Glover and B. Barrett (2010). The complete genome sequence of Canna yellow streak virus. *Archives of Virology* **155**: 1515–1518.

Monstein, H., S. Nikpour-Badr and J. Jonasson (2001). Rapid molecular identification and subtyping of *Helicobacter pylori* by pyrosequencing of the 16S rDNA variable V1 and V3 regions. *FEMS Microbiol Lett* **199**(1): 103-107.

Moreillon, P. and Y. A. Que (2004). Infective endocarditis. *Lancet* **363**(9403): 139-149.

Murray, V. (1989). Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res* **17**(21): 8889.

Naas, T., C. Oxacelay and P. Nordmann (2007). Identification of CTX-M-Type Extended-Spectrum-β-Lactamase Genes Using Real-Time PCR and Pyrosequencing. *Antimicrob Agents Chemother* **51**(1): 223-230.

Naas, T., L. Poirel and P. Nordmann (2006). Pyrosequencing for rapid identification of carbapenem-hydrolysing OXA-type β-lactamases in *Acinetobacter baumannii*. *Clin Microbiol Infect* **12**(12): 1236-1240.

Nakano, K., H. Inaba, R. Nomura, H. Nemoto, M. Takeda, H. Yoshioka, H. Matsue, T. Takahashi, K. Taniguchi, A. Amano and T. Ooshima (2006). Detection of Cariogenic *Streptococcus mutans* in Extirpated Heart Valve and Atheromatous Plaque Specimens. *J Clin Microbiol* **44**(9): 3313-3317.

Neeleman, C., C. H. W. Klaassen, D. M. Klomberg, H. A. de Valk and J. W. Mouton (2004). Pneumolysin Is a Key Factor in Misidentification of Macrolide-Resistant *Streptococcus pneumoniae* and Is a Putative Virulence Factor of *S. mitis* and Other Streptococci. *J Clin Microbiol* **42**(9): 4355-4357.

Nomura, R., K. Nakano, H. Nemoto, K. Fujita, S. Inagaki, T. Takahashi, K. Taniguchi, M. Takeda, H. Yoshioka, A. Amano and T. Ooshima (2006). Isolation and characterization of *Streptococcus mutans* in heart valve and dental plaque specimens from a patient with infective endocarditis. *J Med Microbiol* **55**(Pt 8): 1135-1140.

Norskov-Lauritsen, N. and M. Kilian (2006). Reclassification of *Actinobacillus actinomycetemcomitans, Haemophilus aphrophilus, Haemophilus paraphrophilus* and *Haemophilus segnis* as *Aggregatibacter actinomycetemcomitans* gen. nov., comb. nov., *Aggregatibacter aphrophilus comb.* nov. and *Aggregatibacter segnis* comb. nov., and emended description of *Aggregatibacter aphrophilus* to include V factor-dependent and V

factor-independent isolates. *Int J Syst Evol Microbiol* **56**(9): 2135-2146.

Nyberg, S. D., M. Österblad, A. J. Hakanen, P. Huovinen, J. Jalava and the Finnish Study Group for Antimicrobial Resistance (2007). Detection and Molecular genetics of Extended-Spectrum Beta-Lactamases among cefuroxime- resistant *Escherichia coli* and *Klebsiella* spp. isolates from Finland, 2002-2004. *Scand J Infect Dis* **39**: 417-424.

Nyrén, P. (1987). Enzymatic method for continuous monitoring of DNA polymerase activity. *Anal Biochem* **167**(2): 235-238.

Nyrén, P. (2007). The History of pyrosequencing. Methods in Molecular Biology: Pyrosequencing protocols. S. Marsh. Totowa, New Jersy, USA, Humana Press. **373:** 1-13.

O'Meara, D., K. Wilbe, T. Leitner, B. Hejdeman, J. Albert and J. Lundeberg (2001). Monitoring resistance to human immunodeficiency virus type 1 protease inhibitors by pyrosequencing. *J Clin Microbiol* **39**(2): 464-473.

Parracho, H. M. R. T., M. O. Bingham, G. R. Gibson and A. L. McCartney (2005). Differences between the gut microflora of children with autistic spectrum disorders and that of healthy children. *J Med Microbiol* **54**(10): 987-991.

Paterson, D. L. (2006). Resistance in Gram-Negative Bacteria: *Enterobacteriaceae*. *The American Journal of Medicine* **119**(6, Supplement 1): S20-S28.

Pfeifer, Y., A. Cullik and W. Witte (2010). Resistance to cephalosporins and carbapenems in Gram-negative bacterial pathogens. *Int J Med Microbiol* **300**(6): 371-379.

Picard, F. J., D. Ke, D. K. Boudreau, M. Boissinot, A. Huletsky, D. Richard, M. Ouellette, P. H. Roy and M. G. Bergeron (2004). Use of *tuf* sequences for genus-specific PCR detection and phylogenetic analysis of 28 streptococcal species. *J Clin Microbiol* **42**(8): 3686-3695.

Pihlajamäki, M., T. Kaijalainen, P. Huovinen, J. Jalava and the Finnish Study Group for Antimicrobial Resistance (2002a). Rapid increase in macrolide resistance among penicillin non-susceptible pneumococci in Finland, 1996-2000. *J Antimicrob Chemother* **49**(5): 785-792.

Pihlajamäki, M., J. Kataja, H. Seppälä, J. Elliot, M. Leinonen, P. Huovinen and J. Jalava (2002b). Ribosomal Mutations in *Streptococcus pneumoniae* Clinical Isolates. *Antimicrob Agents Chemother* **46**(3): 654-658.

Pitout, J. D. and K. B. Laupland (2008). Extended-spectrum beta-lactamase-producing *Enterobacteriaceae*: an emerging public-health concern. *Lancet Infect Dis* **8**(3): 159-166.

Poirel, L., T. Naas and P. Nordmann (2006). Pyrosequencing as a Rapid Tool for Identification of GES-Type Extended-Spectrum β-Lactamases. *J Clin Microbiol* **44**(8): 3008-3011.

Pourmand, N., E. Elahi, R. W. Davis and M. Ronaghi (2002). Multiplex Pyrosequencing. *Nucleic Acids Res* **30**(7): e31.

Poyart, C., G. Quesne, S. Coulon, P. Berche and P. Trieu-Cuot (1998). Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *J Clin Microbiol* **36**(1): 41-47.

Poyart, C., G. Quesne and P. Trieu-Cuot (2002). Taxonomic dissection of the *Streptococcus bovis* group by analysis of manganese-dependent superoxide dismutase gene (*sodA*) sequences: reclassification of '*Streptococcus infantarius* subsp. *coli*' as *Streptococcus lutetiensis* sp. nov. and of *Streptococcus bovis* biotype II.2 as *Streptococcus pasteurianus* sp. nov. *Int J Syst Evol Microbiol* **52**(Pt 4): 1247-1255.

Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D. R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.-M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H. B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Dore, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, P. Bork, S. D. Ehrlich and J. Wang (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**(7285): 59-65.

Quince, C., A. Lanzen, T. P. Curtis, R. J. Davenport, N. Hall, I. M. Head, L. F. Read and W. T. Sloan (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Meth* **6**: 639 - 641.

Randegger, C. C. and H. Hächler (2001). Real-Time PCR and Melting Curve Analysis for Reliable and Rapid Detection of SHV Extended-Spectrum β-Lactamases. *Antimicrob Agents Chemother* **45**(6): 1730-1736.

Rappe, M. S. and S. J. Giovannoni (2003). The uncultured microbial majority. *Annu Rev Microbiol* **57**: 369-394.

Rautakorpi, U. M., J. Lumio, P. Huovinen and T. Klaukka (1999). Indication-based use of antimicrobials in Finnish primary health care. Description of a method for data collection and results of its application. *Scand J Prim Health Care* **17**(2): 93-99.

Richter, M. and R. Rosselló-Móra (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U. S. A.* **106**(45): 19126-19131.

Roberts, M. C., J. Sutcliffe, P. Courvalin, L. B. Jensen, J. Rood and H. Seppala (1999). Nomenclature for macrolide and macrolide-lincosamide-streptogramin B resistance determinants. *Antimicrob Agents Chemother* **43**(12): 2823-2830.

Robinson, I., J. Stromley, V. Varel and E. Cato (1988). *Streptococcus intestinalis*, a new species from the *colons* and feces of pigs. *Int J Syst Bacteriol* **38**(3): 245-248.

Ronaghi, M. (2000). Improved Performance of Pyrosequencing Using Single-Stranded DNA-Binding Protein. *Anal Biochem* **286**(2): 282-288.

Ronaghi, M. (2001). Pyrosequencing Sheds Light on DNA Sequencing. *Genome Res* **11**(1): 3-11.

Ronaghi, M. and E. Elahi (2002). Pyrosequencing for microbial typing. *J Chromatogr B Analyt Technol Biomed Life Sci* **782**(1-2): 67-72.

Ronaghi, M., S. Karamohamed, B. Pettersson, M. Uhlen and P. Nyrén (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* **242**(1): 84-89.

Ronaghi, M., M. Nygren, J. Lundeberg and P. Nyrén (1999). Analyses of secondary structures in DNA by pyrosequencing. *Anal Biochem* **267**(1): 65-71.

Ronaghi, M., M. Uhlen and P. Nyrén (1998). A sequencing method based on real-time pyrophosphate. *Science* **281**(5375): 363, 365.

Rudolph, K. M., A. J. Parkinson, C. M. Black and L. W. Mayer (1993). Evaluation of polymerase chain reaction for diagnosis of pneumococcal pneumonia. *J Clin Microbiol* **31**(10): 2661-2666.

Ruoff, K. L., R. Whiley and D. Beighton (2003). *Streptococcus*. Manual of Clinical Microbiology, 8th edition. P. R. Murray, E. J. Baron, M. A. Pfaller, J. H. Jorgensen and R. H. Yolken. Washington, D.C., ASM Press**:** 405-421.

Sanger, F., S. Nicklen and A. R. Coulson (1977). DNA Sequencing with Chain-Terminating Inhibitors. *PNAS* **74**(12): 5463-5467.

Santagati, M., F. Iannelli, C. Cascone, F. Campanile, M. R. Oggioni, S. Stefani and G. Pozzi (2003). The novel conjugative transposon tn1207.3 carries the macrolide efflux gene *mef*(A) in *Streptococcus pyogenes*. *Microb Drug Resist* **9**(3): 243-247.

Santagati, M., F. Iannelli, M. R. Oggioni, S. Stefani and G. Pozzi (2000). Characterization of a Genetic Element Carrying the Macrolide Efflux Gene *mef*(A) in *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* **44**(9): 2585-2587.

Schabereiter-Gurtner, C., A. M. Hirschl, B. Dragosics, P. Hufnagl, S. Puz, Z. Kovach, M. Rotter and A. Makristathis (2004). Novel Real-Time PCR Assay for Detection of *Helicobacter pylori* Infection and Simultaneous Clarithromycin Susceptibility Testing of Stool and Biopsy Specimens. *J Clin Microbiol* **42**(10): 4512-4518.

Schlegel, L., F. Grimont, E. Ageron, P. A. D. Grimont and A. Bouvet (2003). Reappraisal of the taxonomy of the *Streptococcus bovis*/*Streptococcus equinus* complex and related species: description of *Streptococcus gallolyticus* subsp. *gallolyticus* subsp. nov., *S. gallolyticus* subsp. *macedonicus* subsp. nov. and *S. gallolyticus* subsp. *pasteurianus* subsp. nov. *Int J Syst Evol Microbiol* **53**(3): 631-645.

Schlegel, L., F. Grimont, M. D. Collins, B. Regnault, P. A. Grimont and A. Bouvet (2000). *Streptococcus infantarius* sp. nov., *Streptococcus infantarius* subsp. *infantarius* subsp. nov. and *Streptococcus infantarius* subsp. *coli* subsp. nov., isolated from humans and food. *Int J Syst Evol Microbiol* **50 Pt 4**: 1425-1434.

Schleifer, K. H., M. Ehrmann, U. Krusch and H. Neve (1991). Revival of the species *Streptococcus thermophilus* (ex. Orla-Jenson, 1919) nom. rev. *Syst Appl Microbiol* **14**: 386-388.

Schlünzen, F., R. Zarivach, J. Harms, A. Bashan, A. Tocilj, R. Albrecht, A. Yonath and F. Franceschi (2001). Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**(6858): 814-821.

Schouls, L. M., C. S. Schot and J. A. Jacobs (2003). Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. *J. Bacteriol.* **185**(24): 7241-7246.

Schuurman, R., L. Demeter, P. Reichelderfer, J. Tijnagel, T. de Groot and C. Boucher (1999). Worldwide Evaluation of DNA Sequencing Approaches for Identification of Drug Resistance Mutations in the Human Immunodeficiency Virus Type 1 Reverse Transcriptase. *J Clin Microbiol* **37**(7): 2291-2296.

Seng, P., M. Drancourt, F. Gouriet, B. La Scola, P. E. Fournier, J. M. Rolain and D. Raoult (2009). Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Clin Infect Dis* **49**(4): 543-551.

Seppälä, H., A. Nissinen, H. Järvinen, S. Huovinen, T. Henriksson, E. Herva, S. E. Holm, M. Jahkola, M. L. Katila, T. Klaukka and et al. (1992). Resistance to erythromycin in group A streptococci. *N Engl J Med* **326**(5): 292-297.

Seppälä, H., A. Nissinen, Q. Yu and P. Huovinen (1993). Three different phenotypes of erythromycin-resistant *Streptococcus pyogenes* in Finland. *J Antimicrob Chemother* **32**(6): 885-891.

Shames, D. S., J. D. Minna and A. F. Gazdar (2007). Methods for detecting DNA methylation in tumors: From bench to bedside. *Cancer Lett* **251**(2): 187-198.

Shewmaker, P. L., A. C. Camus, T. Bailiff, A. G. Steigerwalt, R. E. Morey and G. Carvalho Mda (2007). *Streptococcus ictaluri* sp. nov., isolated from Channel Catfish Ictalurus punctatus broodstock. *Int J Syst Evol Microbiol* **57**(Pt 7): 1603-1606.

Simmon, K. E., L. Hall, C. W. Woods, F. Marco, J. M. Miro, C. Cabell, B. Hoen, M. Marin, R. Utili, E. Giannitsioti, T. Doco-Lecompte, S. Bradley, S. Mirrett, A. Tambic, S. Ryan, D. Gordon, P. Jones, T. Korman, D. Wray, L. B. Reller, M.-F. Tripodi, P. Plesiat, A. J. Morris, S. Lang, D. R. Murdoch, C. A. Petti and I. and the International Collaboration on Endocarditis Microbiology (2008). Phylogenetic Analysis of Viridans Group Streptococci Causing Endocarditis. *J Clin Microbiol* **46**(9): 3087-3090.

Sinclair, A., C. Arnold and N. Woodford (2003). Rapid detection and estimation by pyrosequencing of 23S rRNA genes with a single nucleotide polymorphism conferring linezolid resistance in Enterococci. *Antimicrob Agents Chemother* **47**(11): 3620-3622.

Sly, L. I., M. M. Cahill, R. Osawa and T. Fujisawa (1997). The tannin-degrading species *Streptococcus gallolyticus* and *Streptococcus caprinus* are subjective synonyms. *Int J Syst Bacteriol* **47**(3): 893-894.

Spellerberg, B. and C. Brandt (2007). *Streptococcus*. Manual of clinical microbiology, 9th Edition. P. R. Murray, J. H. Jorgensen, M. L. Landry and M. A. Pfaller. Washington, D.C., ASM Press**:** 405-421.

Stackebrandt, E., W. Frederiksen, G. M. Garrity, P. Grimont, P. Kampfer, M. Maiden, X. Nesme, R. Rossello-Mora, J. Swings, H. G. Truper, L. Vauterin, A. C. Ward and W. B. Whitman (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**(3): 1043-1047.

Suau, A., R. Bonnet, M. Sutren, J.-J. Godon, G. R. Gibson, M. D. Collins and J. Dore (1999). Direct Analysis of Genes Encoding 16S rRNA from Complex Communities Reveals Many Novel Molecular Species within the Human Gut. *Appl Environ Microbiol* **65**(11): 4799-4807.

Sutcliffe, J., T. Grebe, A. Tait-Kamradt and L. Wondrack (1996). Detection of erythromycin-

resistant determinants by PCR. *Antimicrob Agents Chemother* **40**(11): 2562-2566.

Svantesson, A., P. O. Westermark, J. H. Kotaleski, B. Gharizadeh, A. Lansner and P. Nyrén (2004). A mathematical model of the Pyrosequencing reaction system. *Biophys Chem* **110**(1-2): 129-145.

Tait-Kamradt, A., T. Davies, P. C. Appelbaum, F. Depardieu, P. Courvalin, J. Petitpas, L. Wondrack, A. Walker, M. R. Jacobs and J. Sutcliffe (2000a). Two New Mechanisms of Macrolide Resistance in Clinical Strains of Streptococcus pneumoniae from Eastern Europe and North America. *Antimicrob Agents Chemother* **44**(12): 3395-3401.

Tait-Kamradt, A., T. Davies, M. Cronan, M. R. Jacobs, P. C. Appelbaum and J. Sutcliffe (2000b). Mutations in 23S rRNA and Ribosomal Protein L4 Account for Resistance in Pneumococcal Strains Selected In Vitro by Macrolide Passage. *Antimicrob Agents Chemother* **44**(8): 2118-2125.

Takada, K. and M. Hirasawa (2007). *Streptococcus orisuis* sp. nov., isolated from the pig oral cavity. *Int J Syst Evol Microbiol* **57**(Pt 6): 1272-1275.

Takada, K. and M. Hirasawa (2008). *Streptococcus dentirousetti* sp. nov., isolated from the oral cavities of bats. *Int J Syst Evol Microbiol* **58**(1): 160-163.

Tano, K., E. Grahn-Hakansson, S. E. Holm and S. Hellstrom (2000). Inhibition of OM pathogens by alpha-hemolytic streptococci from healthy children, children with SOM and children with rAOM. *Int J Pediatr Otorhinolaryngol* **56**(3): 185-190.

Tanzer, J. M., J. Livingston and A. M. Thompson (2001). The microbiology of primary dental caries in humans. *J Dent Educ.* **65**(10): 1028-1037.

Teng, L., P. Hsueh, Y. Chen, S. Ho and K. Luh (1998). Antimicrobial susceptibility of viridans group streptococci in Taiwan with an emphasis on the high rates of resistance to penicillin and macrolides in *Streptococcus oralis*. *J Antimicrob Chemother* **41**(6): 621-627.

Teng, L. J., P. R. Hsueh, J. C. Tsai, P. W. Chen, J. C. Hsu, H. C. Lai, C. N. Lee and S. W. Ho (2002). *groESL* sequence determination, phylogenetic analysis, and species differentiation for viridans group streptococci. *J Clin Microbiol* **40**(9): 3172-3178.

Tenson, T., M. Lovmar and M. Ehrenberg (2003). The mechanism of action of macrolides, lincosamides and streptogramin B reveals the nascent peptide exit path in the ribosome. *J Mol Biol* **330**(5): 1005-1014.

Teughels, W., S. Kinder Haake, I. Sliepen, M. Pauwels, J. Van Eldere, J. J. Cassiman and M. Quirynen (2007). Bacteria Interfere with *A. actinomycetemcomitans* Colonization. *J Dent Res* **86**(7): 611-617.

Tindall, B. J., R. Rossello-Mora, H. J. Busse, W. Ludwig and P. Kampfer (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* **60**(1): 249-266.

Tjalsma, H., M. Scholler-Guinard, E. Lasonder, T. J. Ruers, H. L. Willems and D. W. Swinkels (2006). Profiling the humoral immune response in colon cancer patients: diagnostic antigens from Streptococcus bovis. *Int J Cancer* **119**(9): 2127-2135.

Tong, H., X. Gao and X. Dong (2003). *Streptococcus oligofermentans* sp. nov., a novel oral isolate from caries-free humans. *Int J Syst Evol Microbiol* **53**(4): 1101-1104.

Tsakalidou, E., E. Zoidou, B. Pot, L. Wassill, W. Ludwig, L. A. Devriese, G. Kalantzopoulos, K. H. Schleifer and K. Kersters (1998). Identification of streptococci from Greek Kasseri cheese and description of *Streptococcus macedonicus* sp. nov. *Int J Syst Bacteriol* **48**(2): 519-527.

Tuohy, M. J., G. S. Hall, M. Sholtis and G. W. Procop (2005). Pyrosequencing as a tool for the identification of common isolates of *Mycobacterium sp*. *Diagn Microbiol Infect Dis* **51**(4): 245-250.

Turnbaugh, P. J., M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight and J. I. Gordon (2009). A core gut microbiome in obese and lean twins. *Nature* **457**: 480 - 484.

Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight and J. I. Gordon (2007).

The Human Microbiome Project. *Nature* **449**(7164): 804-810.

Täpp, J., M. Thollesson and B. Herrmann (2003). Phylogenetic relationships and genotyping of the genus *Streptococcus* by sequence determination of the RNase P RNA gene, *rnpB*. *Int J Syst Evol Microbiol* **53**(6): 1861-1871.

Unnerstad, H., H. Ericsson, A. Alderborn, W. Tham, M. L. Danielsson-Tham and J. G. Mattsson (2001). Pyrosequencing as a method for grouping of *Listeria monocytogenes* strains on the basis of single-nucleotide polymorphisms in the *inlB* gene. *Appl Environ Microbiol* **67**(11): 5339-5342.

Vacher, S., A. Menard, E. Bernard and F. Megraud (2003). PCR-Restriction Fragment Length Polymorphism Analysis for Detection of Point Mutations Associated with Macrolide Resistance in *Campylobacter* spp. *Antimicrob Agents Chemother* **47**(3): 1125-1128.

Vancanneyt, M., L. A. Devriese, E. M. De Graef, M. Baele, K. Lefebvre, C. Snauwaert, P. Vandamme, J. Swings and F. Haesebrouck (2004). *Streptococcus minor* sp. nov., from faecal samples and tonsils of domestic animals. *Int J Syst Evol Microbiol* **54**(2): 449-452.

Vandamme, P., L. Devriese, F. Haesebrouck and K. Kersters (1999). *Streptococcus intestinalis* Robinson et al. 1988 and *Streptococcus alactolyticus* Farrow et al. 1984 are phenotypically indistinguishable. *Int J Syst Bacteriol* **49**(2): 737-741.

Varaldo, P. E., M. P. Montanari and E. Giovanetti (2009). Genetic Elements Responsible for Erythromycin Resistance in Streptococci. *Antimicrob Agents Chemother* **53**(2): 343-353.

Wayne, L. G. (1988). International Committee on Systematic Bacteriology: announcement of the report of the ad hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Zentralbl Bakteriol Mikrobiol Hyg [A]* **268**(4): 433-434.

Weisblum, B. (1995). Insights into erythromycin action from studies of its activity as inducer of resistance. *Antimicrob Agents Chemother* **39**(4): 797-805.

Vela, A. I., E. Fernandez, P. A. Lawson, M. V. Latre, E. Falsen, L. Dominguez, M. D. Collins and J. F. Fernandez-Garayzabal (2002). *Streptococcus entericus* sp. nov., isolated from cattle intestine. *Int J Syst Evol Microbiol* **52**(2): 665-669.

Vester, B. and S. Douthwaite (2001). Macrolide resistance conferred by base substitutions in 23S rRNA. *Antimicrob Agents Chemother* **45**(1): 1-12.

Westling, K., I. Julander, P. Ljungman, M. Vondracek, B. Wretlind and S. Jalal (2008). Identification of species of viridans group streptococci in clinical blood culture isolates by sequence analysis of the RNase P RNA gene, *rnpB*. *J Infect* **56**(3): 204-210.

Whatmore, A. M., A. Efstratiou, A. P. Pickerill, K. Broughton, G. Woodard, D. Sturgeon, R. George and C. G. Dowson (2000). 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. *Infect Immun* **68**(3): 1374-1382.

Whatmore, A. M. and R. A. Whiley (2002). Re-evaluation of the taxonomic position of *Streptococcus ferus*. *Int J Syst Evol Microbiol* **52**(Pt 5): 1783-1787.

Wheeler, D. A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, Y.-J. Chen, V. Makhijani, G. T. Roth, X. Gomes, K. Tartaro, F. Niazi, C. L. Turcotte, G. P. Irzyk, J. R. Lupski, C. Chinault, X.-z. Song, Y. Liu, Y. Yuan, L. Nazareth, X. Qin, D. M. Muzny, M. Margulies, G. M. Weinstock, R. A. Gibbs and J. M. Rothberg (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**(7189): 872-876.

Whiley, R. A. and D. Beighton (1998). Current classification of the oral streptococci. *Oral Microbiol Immunol* **13**(4): 195-216.

Whiley, R. A., H. Y. Fraser, C. W. Douglas, J. M. Hardie, A. M. Williams and M. D. Collins (1990). *Streptococcus parasanguis* sp. nov., an atypical viridans *Streptococcus* from human clinical specimens. *FEMS Microbiol Lett* **56**(1-2): 115-121.

Willcox, M., H. Zhu and K. Knox (2001). *Streptococcus australis* sp. nov., a novel oral

streptococcus. *Int J Syst Evol Microbiol* **51**(4): 1277-1281.

Williams, A. M. and M. D. Collins (1990). Molecular taxonomic studies on *Streptococcus uberis* types I and II. Description of *Streptococcus parauberis* sp. nov. *J Appl Bacteriol* **68**(5): 485-490.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**(2): 221-271.

Wolff, B. J., W. L. Thacker, S. B. Schwartz and J. M. Winchell (2008). Detection of Macrolide Resistance in *Mycoplasma pneumoniae* by Real-Time PCR and High-Resolution Melt Analysis. *Antimicrob Agents Chemother* **52**(10): 3542-3549.

Woo, P. C., S. K. Lau, J. L. Teng, H. Tse and K. Y. Yuen (2008). Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect* **14**(10): 908-934.

Woo, P. C., D. M. Tam, K. W. Leung, S. K. Lau, J. L. Teng, M. K. Wong and K. Y. Yuen (2002). *Streptococcus sinensis* sp. nov., a novel species isolated from a patient with infective endocarditis. *J Clin Microbiol* **40**(3): 805-810.

Wright, A. J. (1999). The penicillins. *Mayo Clin Proc* **74**: 290-307.

Yarza, P., M. Richter, J. Peplies, J. Euzeby, R. Amann, K.-H. Schleifer, W. Ludwig, F. O. Glöckner and R. Rosselló-Móra (2008). The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* **31**(4): 241-250.

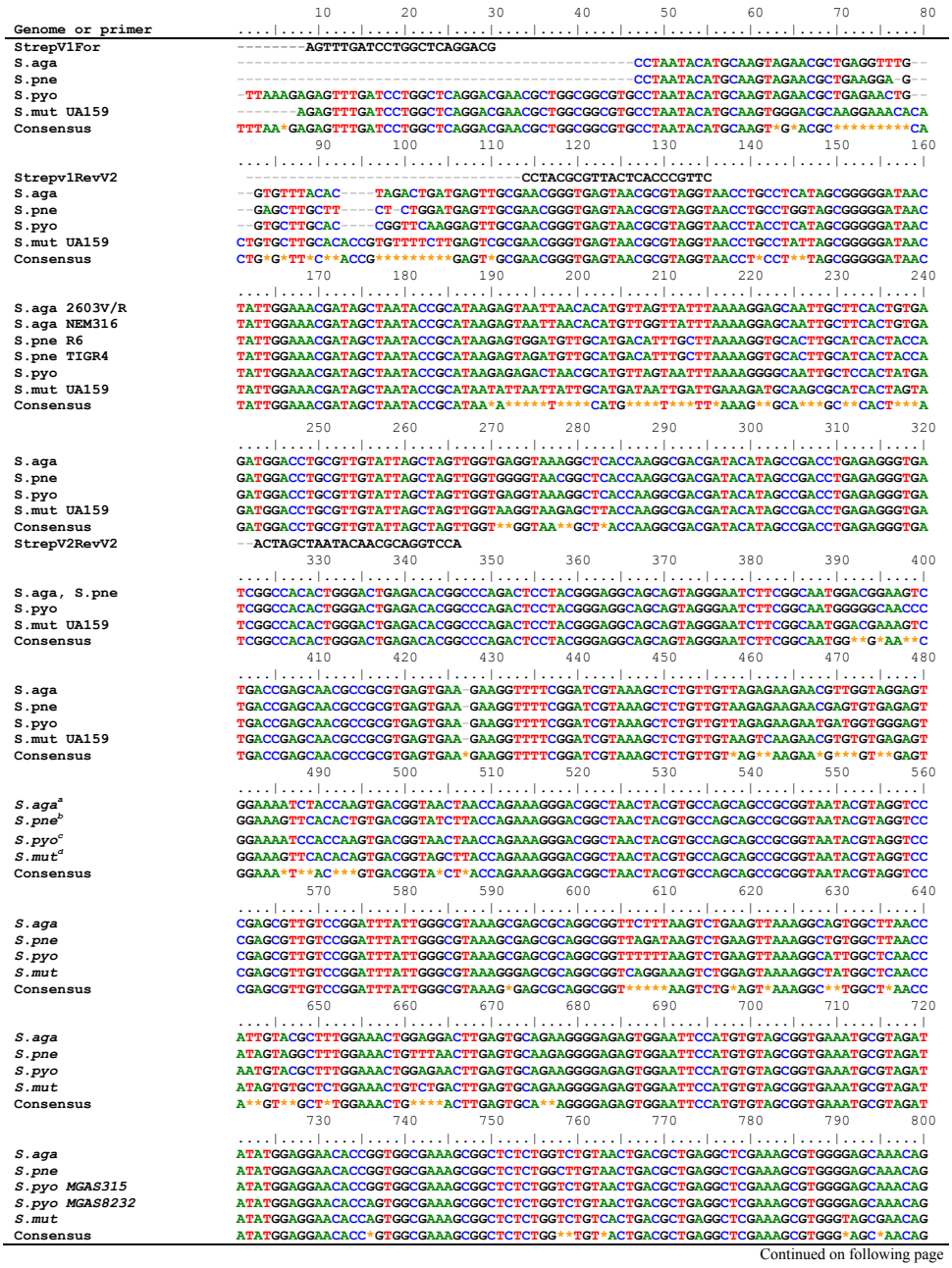
Yonath, A. (2005). Antibiotics targeting ribosomes: Resistance, Selectivity, Synergism, and Cellular Regulation. *Annu Rev Biochem* **74**(1): 649-679.

Zhou, G.-H., M. Gotou, T. Kajiyama and H. Kambara (2005). Multiplex SNP typing by bioluminometric assay coupled with terminator incorporation (BATI). *Nucl Acids Res* **33**(15): e133-.

Zhu, H., M. D. Willcox and K. W. Knox (2000). A new species of oral *Streptococcus* isolated from Sprague-Dawley rats, *Streptococcus orisratti* sp. nov. *Int J Syst Evol Microbiol* **50 Pt 1**: 55-61.

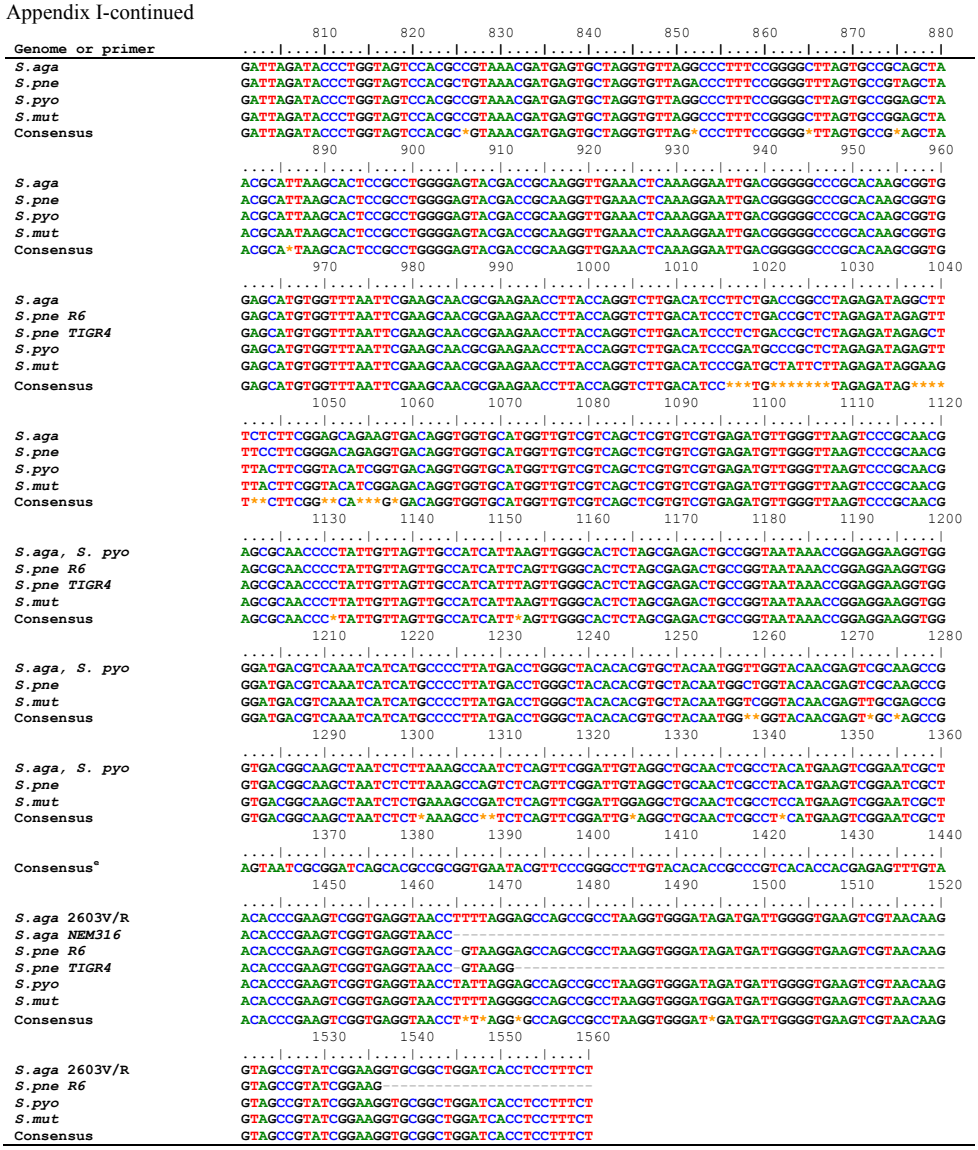
APPENDICES

Appendix I. Alignment of streptococcal genomic sequences based on which the regions used in the pyrosequencing assays were chosen. The sequences and positions of the pyrosequencing primers are also presented.



Continued on following page

Appendix I-continued



[a] *S.aga* = *S. agalactiae*, if no strain designation is given, the sequence was identical in strains 2603/R and NEM316.

[b] *S.pne* = *S. pneumoniae*, if no strain designation is given, the sequence was identical in strains MGAS315, MGAS8232 and M1GAS.

[c] *S.pyo* = *S. pyogenes*, if no strain designation is given, the sequence was identical in strains R6 and TIGR4.

[d] *S.mut* = *S. mutans* UA159

[e] Identical sequence in all genomes

Appendix II. The v1+v2 sequences (n=68) of the 76 streptococcal type strains included in this study.

Species[a]	v1 10 20 30	v2 40 50 60
***S. mitis-S. sanguinis* group**		
S. australis	GCAACTCATCCAGAAAGAGCAAGCTCCTTCTCTGGTAGTGGAGCAATTGCCCCTTTCAAG	
S. infantis	GCAACTCATCCAGAAGAGCAAGCTCCTCCTTCTGGTAGTGGTGCAATTGCACCTTTCAAG	
S. cristatus	GCAACTCATCCAGAAGAGCAAGCTCCTCCTTCTGGTAGTGGTGCATTTGCACCTTTTAAT	
S. pneumoniae	GCAACTCATCCAGAGAAGCAAGCTCCTCCTTCTGGTAGTGATGCAAGTGCACCTTTTAAG	
S. pseudopneumoniae	GCAACTCATCCAGAGAAGCAAGCTCCTCCTTCTGGTAGTGATGCATTTGCACCTTTTAAG	
S. peroris	GCAACTCATCCAGAGAAGCAAGCTCCTCCTTCTGGTAGTGGTGCAATTGCACCTTTTAAA	
S. oralis	GCAACTCATCCGCTCGGTGCAAGCACCAAGTCTGGTAGTGATGCAATTGCACCTTTTAAG	
S. mitis	GCAACTCATCCGGAGAAGCAAGCTCCTCCTTCTGGTAGTGATGCAATTGCACCTTTTAAG	
S. massiliensis	GCAACTCATTAGATTAGTGCAAGCACCAATTCTGGTAGTGATACAGTTGTATCTTTCAAG	
S. orisratti	GCGACTCATGATTAATGGTGGAGCAAGCTCTCTCATAGTGAAGCTCTTGCCCCTTTCAAG	
S. sinensis	GCGACTCACAAAGTACGGTGTAGTAAACTATCTCTTAGTGATGCAATTGCATCTTTTAAT	
S. parasanguinis	GCAACTCCTCCGCTCGGTGCAAGCACCAAGTCTCTTAGTGATGCATTTGCACCTTTCAAG	
S. oligofermentans	GCAACTCTTCCGGAAAGAGCAAGCTCCTTCTCTCTTAGTGATGCATTTGCATCTTTTAAC	
S. entericus	GCGACTCAGTCAATTCGATGGAGCAAGCTCTCTGATAGTGAAGCAATTGCTCCTTTTAAG	
S. gordonii	GCAACTCACAGTCTATGGTGTAGCAAGCTATCTGGTAGTGGTGCAATTGCACCTTTCAAT	
S. sanguinis	GCAACTCATCCAAGAAGAGCAAGCTCCTCTTCTGGTAGTGATGCAATTGCATCTTTCAAT	
***S. anginosus* group**		
S. intermedius	GCAACTCACAGAATACGGTGTAGTAAACTATCTGGTAGTGATGCATTTGCACCTTTTAAA	
S. constellatus[b]	GCAACTCACAGAATACGGTGTAGTAAACTATCTGGTAGTGATGCTCTTGCACCTTTTAAA	
S. anginosus	GCAACTCACAGTCTATGGTGTAGCAAGCTATCTACTAGCGATGCAATTGCATCTTTCAAG	
***S. salivarius* group**		
S. salivarius[c]	GCAACTCATCCAAGAAGAGCAAGCTCCTCTTCTTGTAGTGGAGCAATTGCCCCTTTCAAA	
S. alactolyticus	GCAACTCATCCAAAAGAAGCAAGCTCCTTTTCTCGTAGTGATGCAATTGCATCTTTCAAG	
S. hyointestinalis	GCAACTCATCCAGAAGAAGCAAGCTCCCTCTCTGATAGTGGAGCAGTTGCCCCTTTTAAA	
***S. bovis* group**		
S. lutetiensis[d]	GCAACTCTTCCAACTTTAGCAAGCTAAAGTTCTACTAGTGAAGCAATTGCTCCTTTCAAG	
S. equinus	GCAACTCTTCCAACTTTAGCAAGCTAAAGTTCTACTAGTGAAGCAATTGCTTCTTTCAAG	
S. gallolyticus[e]	GCAACTCCTTCTACTCTAGCAAGCTAAAGTTCTACTAGTGATGCATTTGCATCTTTCAAG	
S. gallolyticus ssp. *pasteurianus*	GCAACTCCTTCTACTCTAGCAAGCTAAAGTTCTACTAGTGATGCAATTGCATCTTTCAAG	
S. gallolyticus ssp. *macedonicus*[f]	GCAACTCTTCCAACTCTAGCAAGCTAAAGTTCTACTAGTGATGCAATTGCATCTTTTAAG	
S. infantarius ssp. *infantarius*	GCAACTCTTCAAACTTTAGCAAGCTAAAGTTCTACTAGTGAAGCAATTGCTCCTTTCAAG	
***S. mutans* group**		
S. criceti	GCGACTCTTCTTCCTTGTGAGTGCAAGCACTCTGATAGTGGTGCCATTGCACCTTTCAAG	
S. downei	GCGACTCATTAATATCAGTGGAGCAAGCTCTCTGATAGTGATACAATGGTATCTTTTAAG	
S. ferus	GCGACTCATGATGAATGGTGGAGCAAGCTCTCTCTTAGCGATACGATTGTATCTTTTAAT	
S. hyovaginalis	GCGACTCATGATTAATGGTGGAGCAAGCTCTCTGATAGTGAAGCAATTGCTCCTTTTAAA	
S. macacae	GCGACGCCGTCCTCCTACGTTAGTACGAACTCTCTTAGCGATACTGTTGTACCTTTCAAT	
S. mutans	GCGACTCAAGAAAACACGGTGTGCAAGCACTCTACTAGTGATGCGCTTGCATCTTTCAAT	
S. ovis	GCGACTCAGTCAATTCGGTGGAGCAAGCTCTCTCGTAGTGGAGCACTTGCTCCTTTCAAA	
S. ratti	GCGACTCAAGAAAACACGATGTGCAAGCACTCTACTAGTGATGCTCTTGCATCTTTCAAG	
S. sobrinus	GCGACTCATTAGTAACACTGGAGCAAGTCCTCTGATAGTGAAGCAATGGCTTCTTTTAAA	
S. orisuis	GCACTCAGCGTTTACGGCGTGGACTACCAGTCTGATAGTGGNGCCATTGCCCCTTTCAAG	
S. devriesei	GCGACTCGTTGTACCGACCATTGTAGCACGTCTCCAGTGTACCGAAGTAACTTCTTATCT	
Beta-haemolytic streptococci		
S. agalactiae[g]	GCAACTCATCAGTCTAGTGTAAACACCAAATCTCACAGTGAAGCAATTGCTCCTTTTAAA	
S. canis	GCAACTCCTTAGACTAGTGCAAGCACCTGTTCTCATAGTGGAGCAGTTGCCCCTTTTAAG	
S. didelphis	GCAACTCCTTAGACTAGTGCAAGCACCAGTTCTCATAGTGGAGCATTTGCCCCTTTTAAA	
S. dysgalactiae ssp. *dysgalactiae*	GCAACTCCTTGAACCGGTGCAAGCACCAGTTCTCATAGTGATGCAGTTGCACCTTTTAAA	
S. dysgalactiae ssp. *equisimilis*	GCAACTCCTTGGACCGGTGCAAGCACCAGTTCTCATAGTGATGCAGTTGCACCTTTTAAA	
S. equi ssp. *zooepidemicus*	GCGACTCACAGATAATTGTAGCAAGCTACGTCTCATAGTGGAGCTGTTGCCCCTTTTAAA	
S. equi ssp. *equi*	GCGACTCACAGATAATTGTAGCAAGCTACGTCTCATAGTGGAGCTGTTGCTCCTTTTAAA	
S. equi ssp. *ruminatorum*	GCAACTCATCAGTCTAGTGCAAGCACNNNNTCTCATAGTGAAGCGTTTGCTTCTTTCAAG	
S. iniae	GCAACTCTTTGGATTAGTGCAAGCACCAATTCTCATAGTGAAGCATTTGCTCCTTTTAAC	
S. phocae	GCAACTCCTCTTACTTAGTGCAAGCACTAATCTCATAGTGGAGCAATTGCCCCTTTTAAA	
S. porcinus	GCAACTCATTAGACTGGTGCAAGCACCTGTTCTCATAGTGGAGCAGTTGCCCCTTTTAAG	
S. pseudoporcinus	GCAACTCCTTGGTCTAGTGCAAGCACCAGATCTCATAGTGAAGCTTTTGCTCCTTTTAAC	
S. pyogenes	GCAACTCCTTGAACCGGTGCAAGCACCAGTTCTCATAGTGGAGCAATTGCCCCTTTTAAA	
Ungrouped streptococci		
S. acidominimus	GCAACTCCTCAACTCTGTACAAGTACAAAGTCTGGTAGTGGTGCAGTTGCACCTTTTAAA	
S. parauberis	GCAACTCATCTGACTAGTGCAAGCACCAGTTCTCATAGTGAAGCAATTGCTCCTTTTAAA	
S. pleomorphus	GCCACTGACTTCCAAAAAAAGCAAGCTTTTTCCATGTTCAGGGCCGTAGCCCCTTTAATA	
S. pluranimalium	GCGACTCAATCAATTCAAATGTAGCAAGCTTCTCATAGTGAAGCTGTTGCTCCTTTTAAA	
S. thoraltensis	GCGACTCATTTATACCAATTGTAGCAAGCTTCTTTAAGTGATGCTGTTGCACCTTTTAAA	
S. uberis	GCAACTCCTTAGACCAGTGCAAGCACCAGTTCTCATAGTGAAGCATTTGCCCCTTTTAAA	
S. urinalis	GCAACTCCTTGGTCTTGTGCAAGCACATGATCTCATAGTGAAGCAGTTGCTCCTTTTAAG	
S. gallinaceus	GCAACTCATCCGCTCGGTGCAAGCACCAAGTCTGGTAGTGAAGCAATTGCTCCTTTCAAA	
S. castoreus	GCAACTCCTCCAGACTAGTGCAAGCACCAGTCTCTTAGTGAAACTTTTGTTCCTTTCAAC	
S. halichoeri	GCAACTCCTTGAACTAGTGCAAGCACCAGTTCTCATAGTGGAGCTGTTGCCCCTTTTAAG	
S. ictaluri	GCACTCATCGTTTACGGCGTGGACTACCAGTCTCATAGTGAAGCAATTGCCCCTTTTAAA	
S. minor	GCACTCATCGTTTACGGCGTGGACTACCAGTCTCTTAGTGAAGCAGTTGCTCCTTTTAAA	
S. marimammalium	GCACTCATCGTTTACGGCGTGGACTACCAGTCTTCTAGTGATACAATTGTATCTTTTAAC	
S. caballi	GCGACTCATGATTAATGGTGGAGCAAGCTCTCTCGTAGTGATGCTGTTGCATCTTTCAAG	
S. henryi	GCGACTCATGATTAATGGTGTAGCAAGCTATCTACTAGTGAAGCATTTGCTCCTTTTAAA	
S. suis	GCAACTCATCCGTCTAGTGCAAGCACCAGATCTCATAGTGAAGCAATTGCTCCTTTCAAA	

Explanations of the superscripts on following page.

[a] The type strain designations are presented in Table 2.
[b] *S. constellatus* ssp. *constellatus* and *S. constellatus* ssp. *pharynges*
[c] *S. salivarius, S. vestibularis,* and *S. thermophilus*
[d] *S. lutetiensis, S. infantarius* ssp. *coli,* and *S. bovis*
[e] *S. gallolyticus* ssp. *gallolyticus* and *S. caprinus*
[f] *S. gallolyticus* ssp. *macedonicus* and *S. waius*
[g] *S. agalactiae* and *S. difficilis*