Accuracy of using the lytA gene to distinguish Streptococcus pneumoniae from related species

Thomas Greve and Jens Kjølseth Møller†

Department of Clinical Microbiology, Aarhus University Hospital, Skejby, Brendstrupgaardsvej 100, 8200 Aarhus N, Denmark

The need for a microbial identification of Streptococcus pneumoniae independent of culture methods has resulted in the introduction of other laboratory principles. The verification of a proper and exclusive gene for the detection of the pneumococcus by nucleic acid-based tests is, however, still unresolved. A previously published lytA-gene-specific real-time PCR method was applied to a panel of bacterial strains to clarify the analytical sensitivity and specificity of a PCR assay targeting this gene. Furthermore, a phylogenetic analysis of published lytA gene sequences was performed to look at gene sequence differences and the theoretical match with the primers and probes. The lytA-gene-specific PCR detected 46/46 S. pneumoniae isolates. All 49 of the non-pneumococcal isolates tested negative, including 22 isolates from the mitis group streptococci. Phylogenetic analysis of 94 sequences of the lytA gene from different strains of S. pneumoniae, Streptococcus mitis and Streptococcus pseudopneumoniae showed that 70/87 S. pneumoniae sequences constituted one cluster and a further six sequences were outside but adjacent to this cluster, all with a complete match with primers and probes. The remaining 11 S. pneumoniae strains could be placed in a different cluster, which also contained the five S. mitis and two S. pseudopneumoniae strains. All strains had no match with primers and probes. The S. pneumoniae strains in the second cluster were all characterized by being bile-insoluble, an infrequent pneumococcal phenotype. Routine laboratories can utilize the additional observation that pneumococci that were negative by the specific PCR also carried the phenotype of bile insolubility, thereby observing the incidence of false-negative results produced by the PCR assay. The real-time PCR targeting the lytA gene thus constitutes a sensitive and specific assay that distinguishes S. pneumoniae from its close relatives in the mitis group.

INTRODUCTION

Various microbiological methods are used to detect Streptococcus pneumoniae. The analytical and clinical sensitivity and specificity of tests vary. Some methods are entirely dependent on bacterial viability, whereas others are independent. Nucleic acid-based tests are applied frequently as an example of a method that is independent of bacterium viability. These tests are also characterized by a high degree of specificity. Despite this, the verification of a proper and exclusive gene for the detection of the pneumococcus by nucleic acid-based tests is still unresolved.

The mitis group streptococci are thought to be a product of an evolution from a primary ancestral bacterium adapted to one host. The host-species expansion through an evolutionary bottleneck resulted in great diversity (Donati et al., 2010; Kilian et al., 2008). This is detectable at the gene level, by molecular characterization, and often morphologically through phenotypic characterization. The questions raised by this are: (1) do significant sequence differences exist in the lytA gene within the mitis group streptococci? (2) can these differences contribute to the reliable detection of S. pneumoniae?

The use of phy and lytA genes alone or in combination with each other for the detection of S. pneumoniae has been examined previously (Carvalho et al., 2007; Corless et al., 2001; Kee et al., 2008; Sheppard et al., 2004; Strålin et al., 2005). Alternative genes have also been proposed (Carvalho et al., 2007; Kilian et al., 2008; Suzuki et al., 2005, 2006; Zbinden et al., 2011). The main problem in the identification of S. pneumoniae, however, is its striking similarity to the four commensal species Streptococcus mitis, Streptococcus oralis, Streptococcus infantis and Streptococcus pseudopneumoniae (Arbique et al., 2004; Carvalho et al., 2003; Hanage et al., 2005; Kilian et al., 2008; Whatmore et al., 2000).

We present a study applying a lytA-gene-specific real-time PCR method, previously published by Sheppard et al.
(2004), to a panel of bacterial strains to clarify the analytical sensitivity and specificity of a PCR assay targeting this gene. Furthermore, a phylogenetic analysis of published lytA gene sequences was performed to look at gene sequence differences and the theoretical match to the primers and probes in these sequences.

**METHODS**

**Strains and sequences.** A collection of 46 pneumococcal isolates and 49 other isolates were used for the sensitivity and specificity panels. For a phylogenetic analysis, we used 94 partial sequences of the lytA gene composed of 67, 5 and 2 sequences from *S. pneumoniae*, *S. mitis* and *S. pseudopneumoniae*, respectively.

**Bacterial cultures.** *S. pneumoniae* strain ATCC 6301 was used as a positive control for the assays. To verify that the assay could detect a variety of *S. pneumoniae* serotypes and strains, a panel of clinical isolates composed of 23 strains (representing the common vaccine serotypes) of *S. pneumoniae* (serotyped by the Neisseria and *Streptococcus* Reference Laboratory at Statens Serum Institut, Copenhagen, Denmark) was used. Furthermore, 16 less-common serotypes combined with the ATCC 49619 strain and five SK strains from the Kilian Streptococcol Collection (SK618, SK676 = CCUG 28588, SK862, SK865 and SK867) were added. A specificity panel consisting of 49 isolates, including 15 species of streptococci represented by 29 strains (of which 22 were from the *mitis* group), and examples of 18 other closely related species or examples of common pathogens (represented by 20 strains), was assembled from the ATCC, SK, the NCTC or CCUG collections. The *mitis* group streptococci (except *Streptococcus sanguinis*, SK1778) were blinded to us during the study until the data analysis phase. Information about the strains can be found in Supplementary Table S1 in JMM Online.

**Extraction of DNA from bacterial cultures.** The bacterial strains were cultured, after which the colonies were suspended in autoclaved diethylpyrocarbonate-treated water. DNA was extracted from the solution using either the bioMérieux EasyMag robot or the Roche MagNa Pure Compact robot, both according to the manufacturer’s instructions.

**Autolysin primers and probe design.** The primers and probes developed by Sheppard et al. (2004) were used. The primer and probe sites chosen in 2004 showed complete agreement with the 20 *S. pneumoniae* strain sequences that were available from GenBank at the time, but not with sequences from *S. mitis* and *S. oralis* strains reported containing the autolysin gene (Whatmore et al., 2000).

**Internal process control (IPC) primers and probe design.** Murine cytomegalovirus was chosen as an IPC. Primers and probes targeting within the m05 gene (GenBank accession no. U68299.1) were designed by TIB MOLBIOL (Berlin, Germany). The primers, m05F: 5′-CCACTCTAAAGGGCGGCAAC-3′ and m05R: 5′-CGGACCTATCGG-CGTACAATC-3′, amplified a 295 bp fragment. The IPC product was detected by a second set of hybridization probes, m05-FL5′- CGCAAGCGTCATGGGCGC–FL and m05-LC5′-LC705-TAGAATGGTGCTTGGAATGGGACAC–PH, specific to the amplified sequence within the m05 gene.

**LightCycler autolysin PCR.** The protocol was a modification of the one suggested by Sheppard et al. (2004). The main modifications were: the use of murine cytomegalovirus as an IPC; change in the stock concentration of reagents; and a shift to 50 cycles of denaturation. Each isolate was tested in duplicate. An aliquot (5 μL) of the extracted DNA sample was added to the reaction capillary (Roche Diagnostics) containing 15 μL of the reaction mixture. Each LightCycler run contained a negative control and purified pneumococcal DNA as a positive control. Samples were recorded as positive where a crossing point of fewer than 50 cycles was noted. Additionally, the melting temperature peak of the probe, revealed by the melting curve analysis, had to be above 60 °C.

**Selection of sequences.** The theoretical work consisted of a lytA gene sequence analysis of orthologous gene clusters within the *mitis* group. The Protein/Cluster search on the Sybil homepage (Strepneumo, accessed 2011; http://strepneumo-sybil.igs.umaryland.edu/cgi-bin/current/shared/index.cgi?site=strepneumo) was used to identify sequences of the autolysin gene of the TIGR4 *S. pneumoniae* and all the 59 sequences of the orthologous cluster of the 35 strains available on that site. The 957 bp sequence of SP_1937 from TIGR4 was used in a nucleotide BLAST using the National Center for Biotechnology Information (NCBI) homepage. A nucleotide search using MEGA BLAST in the nucleotide collection with a restriction to the five organisms *S. pneumoniae* (taxid: 1313), *S. mitis* (taxid: 28037), *S. pseudopneumoniae* (taxid: 257758), *S. oralis* (taxid: 1303) and *S. infantis* (taxid: 68892) found 91 sequences. Sixty-nine sequences had a query coverage of 92 % or more and a maximum identity of 82–100 %. Two *mitis*-contig sequences were provided by M. Kilian (Department of Biomedicine – Medical Microbiology and Immunology). A total of 152 sequences from 106 strains were selected. No match was found within *S. oralis* and *S. infantis*.

**Adaptation of the lytA gene sequences.** Removal of duplicates from different search engines and use of the query coverage and maximum identity criteria reduced the 152 sequences to 118 sequences in 90 strains, which were aligned using the MEGA version 4.0 software (Supplementary Table S2) (Tamura et al., 2007). The sequences of two of the strains of *S. pneumoniae*, SP9 and S3, were too short and, therefore, were not used in further analysis. A further reduction was made by removal of duplicates/triplicates within a strain, where the gene name and the function of the gene did not match those of the lytA gene. A total of 94 sequences of 693 bp in 88 different strains were left for phylogenetic analysis.

**Phylogeny.** MEGA version 4.0 was used to construct the phylogenetic trees. Maximum parsimony with the use of all sites was applied. A tree was constructed (Fig. 1) showing the relation between the sequences from *S. pneumoniae*, *S. mitis* and *S. pseudopneumoniae* with the gene duplicates/triplicates, and the sequences of the 11 bile-insoluble pneumococci. Bootstrap values (%) were based on 1000 replications (only values above 70 % are shown).

**RESULTS AND DISCUSSION**

**Specific lytA gene PCR, sensitivity and specificity.** The PCR tested positive on 46/46 *S. pneumoniae* strains. All non-*S. pneumoniae* strains (0/49) were found to be test-negative, including the 22 *mitis* group strains. The SK676 strain, which primarily tested negative, is identical to the type strain of *S. pneumoniae* (taxid: 1313; ATCC 33400; NCTC 7465; CCUG 28588). CCUG 28588 was ordered and tested positive. The discrepancy was probably caused by mislabelling of the SK676 strain.

Seventeen of the isolates of the *mitis* group strains had already been characterized (Kilian et al., 2008). The presence of the lytA gene in these strains was based on this publication; depicted in Supplementary Table S1 in JMM Online. The ability to distinguish *S. pneumoniae* from...
S. mitis and S. pseudopneumoniae by the applied PCR is supported by the data on the lytA gene in the study of Kilian et al. (2008).

**Sequence analysis, sensitivity and specificity**

Blasting the sequence of the lytA gene confirmed that the gene is present in many of the mitis group strains, as described by several other authors (Carvalho et al., 2007; Kilian et al., 2008; Llull et al., 2006; Whatmore et al., 2000). We found the gene in S. pneumoniae to be different in general from the lytA gene in other mitis group streptococci (shown in Fig. 1). This has previously been described in the literature (Llull et al., 2006). All 70 S. pneumoniae sequences found in the S. pneumoniae cluster had a complete match with the primers and probes. All of the completely sequenced strains of S. pneumoniae (13) and S. mitis (1) available at NCBI’s homepage were represented in these 70 sequences. The search found no complete genomes of S. pseudopneumoniae or S. infantis, whereas one genome of S. oralis was available, but with no match for the 957 bp lytA gene sequence, as already mentioned. Four [670-6B, JJA, P1031 (complete genomes) and SP14-B569 (draft genome)] and one (CDC305906, draft genome) of the pneumococcal strains had duplicate or triplicate sequences, respectively. All had a complete match with the primers and probes. It is hard to establish whether the six sequences are found within the genomes. The finding could also be a result of mismatch assembly of contigs. Determining this was, however, not the aim of this study.

As shown in Fig. 1, there were slight differences in the gene sequences in the five strains with duplicate/triplicate copies. One copy was found in the main pneumococcal cluster and the other copy was outside. Eleven bile-insoluble S. pneumoniae strains that have been previously studied (Obregon et al., 2002) were found in another cluster with the seven mitis/pseudopneumoniae streptococci, all with an incomplete match with the primers and probes. There was a similar degree of incomplete match with both primers and probes, as shown in Supplementary Table S2. The high specificity of the test could be due to the assay match with either the primers and the probes or just the probes. This was not examined further. Gel electrophoresis could have detected a primer-based amplification product but with no match of the probes.

**Clinical reliability of the lytA gene as a PCR target**

The combined data of the present study showed that the target within the lytA gene is specific to S. pneumoniae with the exception of bile-insoluble pneumococci. Eleven of 133 strains and sequences of pneumococci (46 + 87) were missed by the primers and probes. All of these strains were bile-insoluble, a phenotype found in under 2% of clinical isolates (Richter et al., 2008). However, the pneumococcal strains sequenced and submitted to open access
Species distinction in the mitis group by lytA
databases are not necessarily representative of the majority of clinical isolates. The study found that none of the 29 strains of the close relatives of the pneumococci were positive by either test or sequence analysis. The bile solubility test specifically tests the functionality of the autolysin (Mossor & Tomasz, 1970). This enzyme is a product of gene expression when the lytA gene is transcribed (Garcia et al., 1994). Alterations of the lytA gene sequence can affect the bile solubility test, and might affect the outcome of the PCR. The combination of the sequence differences and the bile-insoluble phenotype can then be explained (Garcia et al., 1994). This finding can be utilized by routine microbiological laboratories as a surrogate marker for the incidence of false-negative results produced by the lytA-gene-specific PCR assay.

This study supports previous results that show that the lytA gene is present in S. pneumoniae and its close relatives. However, we found that differences in the sequences of the lytA gene are greater between the species than within. Ongoing modification of the PCR assay according to these differences may retain the sensitivity and specificity of the ‘pneumococcal’ lytA gene as indicated in a previous study (Carvalho et al., 2007).

Conclusions
The method published by Sheppard et al. (2004) targeting the lytA gene constitutes a sensitive and specific assay for distinguishing S. pneumoniae from its close relatives in the mitis group. This is due to differences in the lytA gene sequence of S. pneumoniae and the other mitis group streptococci. This study suggests that the bile-insoluble pneumococcal strains test negative in the lytA gene PCR, but since this phenotype is found in under 2% of clinical isolates, the assay is applicable in the routine laboratory. However, we recommend that the bile solubility test is regularly performed on pneumococcal isolates. This gives an estimate of the incidence of the phenotype. Furthermore, we recommend validating a lytA-specific PCR by continuous monitoring of published sequences of the lytA gene, and corroborating the finding that bile-insoluble pneumococcal strains are more likely to be found outside the typical S. pneumoniae lytA gene cluster.

Acknowledgements
The authors would like to thank Professor Mogens Kilian for providing the streptococcal strains, helping with the sequence analysis and for the very constructive comments on the manuscript. We would also like to thank Lotte Lambertsen, Manager of the Neisseria and Streptococcus Reference Laboratory at the Statens Serum Institut, for providing the following streptococcal serotypes: 2, 7A, 8, 9A, 10A, 17F, 23F and 35B.

References


