In silico assessment of virulence factors in strains of Streptococcus oralis and Streptococcus mitis isolated from patients with Infective Endocarditis

Rasmussen, Louise H.; Iversen, Katrine Højholt; Dargis, Rimtas; Christensen, Jens Jørgen; Skovgaard, Ole; Justesen, Ulrik Stenz; Rosenvinge, Flemming S.; Moser, Claus; Lukjancenko, Oksana; Rasmussen, Simon

Published in:
Journal of Medical Microbiology

Link to article, DOI:
10.1099/jmm.0.000573

Publication date:
2017

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
In silico assessment of virulence factors in strains of *Streptococcus mitis* and *Streptococcus oralis* isolated from patients with Infective Endocarditis.

Authors:
Louise H. Rasmussen¹,²*, Katrine Højholt¹,⁷*, Rimtas Dargis¹, Jens Jørgen Christensen¹,⁸, Ole Skovgaard², Ulrik S. Justesen³,⁵, Flemming S. Rosenvinge⁴, Claus Moser⁵, Oksana Lukjancenko⁶, Simon Rasmussen⁷ & Xiaohui C. Nielsen¹

*Shared first authorship. Louise H. Rasmussen and Katrine Højholt contributed equally

Louise Hesselbjerg Rasmussen¹,²  lohra@regionsjaelland.dk
Katrine Højholt¹,⁷  katrine@cbs.dtu.dk
Rimtas Dargis¹  rida@regionsjaelland.dk
Jens Jørgen Christensen¹,⁸  jejc@regionsjaelland.dk
Ole Skovgaard²  olesk@ruc.dk
Ulrik Stenz Justesen³,⁵  ujustesen@health.sdu.dk
Flemming Schønning Rosenvinge⁴  flemming.rosenvinge@rsyd.dk
Claus Moser⁵  moser@dadlnet.dk
Oksana Lukjancenko⁶  oklu@food.dtu.dk
Simon Rasmussen⁷  simon@cbs.dtu.dk

**Corresponding author:** Xiaohui Chen Nielsen¹  xcn@regionssjaelland.dk

Affiliations and addresses:

¹The GenBank accession numbers for the 40 genomes are available through the Bioproject accession number PRJNA304678
Abstract

Purpose. *Streptococcus oralis* and *Streptococcus mitis* belong to the Mitis group, which are mostly commensals in the human oral cavity. Even though *S. oralis* and *S. mitis* are oral commensals, they can be opportunistic pathogens causing infective endocarditis. A recent taxonomic re-evaluation of the Mitis group has embedded the species *Streptococcus tigurinus* and *Streptococcus dentisani* into the species *S. oralis* as subspecies. In this study, the distribution of virulence factors that contributes to bacterial immune evasion, colonisation and adhesion were assessed in clinical strains of *S. oralis* (subsp. *oralis*, subsp. *tigurinus* and subsp. *dentisani*) and *S. mitis*.

Methodology. Forty clinical *S. oralis* (subsp. *oralis*, *dentisani* and *tigurinus*) and *S. mitis* genomes were annotated with the pipeline PanFunPro and aligned against the VFDB database for assessment of virulence factors.

Results/Key findings. Three homologs of *pavA, psaA* and *lmb*, encoding adhesion proteins, were present in all strains. Seven homologs of *nanA, nanB, ply, lytA, lytB, lytC* and *iga* with importance for survival in blood and modulation of the human immune system were variously present in the genomes. Few *S. oralis* subspecies specific differences were observed. *iga* homologs were identified in *S. oralis* subsp. *oralis* whereas *lytA* homologs were identified in *S. oralis* subsp. *oralis* and subsp. *tigurinus*.

Conclusion. Differences in presence of virulence factors between the three *S. oralis* subspecies were observed. The virulence gene profiles of the 40 *S. mitis* and *S. oralis* (subsp. *oralis*, subsp. *dentisani* and subsp. *tigurinus*) contribute with important knowledge of these species and new subspecies.
Keywords: Mitis group streptococci - Comparative genomics - Virulence factors - Infective Endocarditis - *Streptococcus mitis* - *Streptococcus oralis*.
**Introduction**

*Streptococcus oralis* and *Streptococcus mitis* are non-hemolytic streptococci belonging to the Mitis group, which mostly are commensals in the human oral cavity throughout life [1, 2]. Even though *S. oralis* and *S. mitis* are oral commensals, they can be opportunistic pathogens entering the bloodstream and causing infective endocarditis (IE) [3, 4]. *Streptococcus tigurinus* and *Streptococcus dentisani* are other members of the Mitis group that have likewise been isolated from the oral cavities [5, 6]. *S. tigurinus* has been described as an IE causing agent [7]. A recently taxonomic re-evaluation of the Mitis group has embedded the two newer species *Streptococcus tigurinus* and *Streptococcus dentisani* as subspecies into the species *S. oralis* [8]. Today the species *S. oralis* consist of the three subspecies *S. oralis* subsp. *oralis*, *S. oralis* subsp. *tigurinus* and *S. oralis* subsp. *dentisani* [8].

*Streptococcus pneumoniae*, another member of the Mitis group, is the closest relative to *S. oralis* and *S. mitis*. Besides colonising the human nasopharynx, *S. pneumoniae* also causes local infections and serious life-threatening diseases, such as septicaemia, meningitis, pneumonia and more rare IE [9-11]. Virulence genes contributing to colonisation (e.g. *nanA, nanB, lytA, lytB, lytC*, and *ply*), contributing to evasion of the immune system (e.g. *iga, cps*) and contributing to adhesion (e.g. *psaA* and *pavA*) have been discovered in *S. pneumoniae* [12-20]. In addition, many of these genes have been identified in *S. mitis* and *S. oralis*.

The Immunoglobulin A1 (IgA1) protease has been observed in both *S. oralis* and *S. mitis*, though variously present in both species [8, 21]. The gene encoding the pneumococcal surface adhesion A (*psaA*) has been identified in all investigated *S. mitis* and *S. oralis* [22, 23] and horizontal *psaA* gene transfer has been suggested among the species in the Mitis group [23]. The genes *ply* and *lytA* have both been recognized in the genomes of a minority of *S. mitis* genomes, but not in the genomes of *S. oralis* [24, 25]. In contrast, both *S. mitis* and *S. oralis* exhibit neuraminidase activity when grown in Brain Heart Infusion broth [26]. A widespread presence of the gene *pavA* was observed in a study
where all nine included *S. mitis* and 11 *S. oralis* strains hybridized with *pavA* illustrating the importance of adherence and virulence protein A (PavA) for oral streptococci [25].

Studies of virulence factors in clinical strains of *S. mitis* and *S. oralis* subsp. *oralis*, subsp. *tigurinus* and subsp. *dentisani* have been limited. We have previously whole genome sequenced and identified 40 *S. mitis* and *S. oralis* isolated from patients with IE [27]. In this study, we identify virulence factors in these *S. mitis* and *S. oralis* genomes in order to identify the distribution of virulence genes with importance for immune evasion, colonisation and adhesion in *S. mitis*, *S. oralis* subsp. *oralis*, *S. oralis* subsp. *dentisani* and *S. oralis* subsp. *tigurinus*.

**Materials and methods**

**Bacterial strains**

Forty blood culture strains, *S. mitis* (*n*=12), *S. oralis* subsp. *oralis* (*n*=14), *S. oralis* subsp. *tigurinus* (*n*=8) and *S. oralis* subsp. *dentisani* (*n*=6) from patients with verified IE were collected retrospectively (2006-2013) from the Capital Region of Denmark (RH strains), Region Zealand (AE, Y and B strains) and Region of Southern Denmark (OD strains). One strain per patient was included in the study, except for one patient who contributed with two strains (B007274_11 and Y11577_11). The verification of IE was conducted by cardiologist and microbiologist according to the modified Duke criteria [28]. The 40 strains had been paired-end sequenced with 100X coverage using Illumina HiSeq 2000 (BGI-Tech Solutions, Hong Kong, China) [27]. The draft genomes were *de novo* assembled with SPAdes [29]. The species identification was based on Multi Locus Sequence Analysis (MLSA), and core-genome phylogeny [8, 27]. The GenBank accession numbers for the 40 genomes are available through the Bioproject accession number PRJNA304678.
Genome annotation

The pipeline PAN-genome analysis based on FUNctional PROfiles (PanFunPro) [30] was used for gene prediction and for prediction of functional domains in the de novo assembled genomes. First genes were predicted and translated into protein sequences using prodigal v2.50 [31]. The translated protein sequences for each streptococcal genome were searched against the databases; PfamA [32], TIGRFAM [33] and SUPERFAMILY [34] using InterProScan software [35] for prediction of functional domains. The combination of non-overlapping functional domains in the protein sequences constituted the functional profiles. Each functional profile was based on a coding sequence.

Hierarchical clustering of species

A presence-absence gene matrix based on the pan-genome of 40 clinical S. mitis and S. oralis strains was constructed in order to get an impression of co-existing genes among the strains examined from the two species. The matrix was constructed using PanGenome2Abundance.pl in PanFunPro [30]. The Pearson correlation coefficient between the 40 strains using their presence/absence functional profiles were basis for hierarchical clustering of the strains.

Prediction of putative virulence genes

Basic Local Alignment Search Tool (BLASTP) [36] was applied to search the translated protein sequences against Virulence Factors of Pathogenic Bacteria database (VFDB), (Accessed 25 August 2015) which contains various virulence factors from other streptococci, Staphylococcus aureus and Enterococcus faecalis [37-39]. The threshold for hits were an e-value < 0.001, a bit score > 50 and a sequence identity percent > 40%. The best hit was based on highest bit score.
Results

Whole genome sequence characterisation

The number of scaffolds from the de novo assembly ranged from 17-85 (S. mitis), 20-41 (S. oralis subsp. dentisani), 7-47 (S. oralis subsp. oralis) and 7-47 (S. oralis subsp. tigurinus). The estimated sizes of the S. mitis and S. oralis genomes ranged from 1.8 Mb-2.1 Mb. Each functional profile was considered based on a coding sequence. Between 1,692-2,083 functional profiles were predicted in the 12 S. mitis strains and 1,734-2,035 functional profiles were predicted in the 28 S. oralis strains. There was no subspecies specific differences between the number of functional profiles in the 28 S. oralis strains. The GC content was slightly higher in S. oralis (40.75-41.50 %) than in S. mitis (39.71-40.28 %). Number of scaffolds, N50, the longest sequences and the number of functional profiles in the 40 S. mitis and S. oralis genomes are presented in Appendix A.

When clustering the strains based on presence/absence of the functional profiles, a tight cluster containing the S. mitis were identified (Fig. 1). The S. oralis strains clustered into three subclusters, which were congruent with earlier observed subclusters based on core-gene phylogeny [27]. Furthermore, the subclustering of S. oralis were congruent with the division of the strains into the three subspecies S. oralis subsp. oralis, subsp. tigurinus and subsp. dentisani [8]. Two S. oralis strains (S. oralis B007274_11 and S. oralis Y11577_11) with high correlation were isolated from the same patient within a day and should be considered as the same strain.

Virulence genes present in S. mitis and S. oralis subsp. oralis, subsp. tigurinus and subsp. dentisani.

In order to determine the presence of virulence genes in S. mitis and S. oralis subsp. oralis, subsp. tigurinus and subsp. dentisani, the functional profiles based on coding sequences in the 40 strains
were aligned against the VFDB database. The number of strains that contained the putative
virulence genes and the protein sequence identity to the VFDB reference sequence are specified in
Table 1. Genes encoding proteins homologous to Adherence and virulence protein A (PavA)
Laminin binding protein (Lmb) and Pneumococcal surface adhesion A (PsaA) were identified in all
40 strains.

Homologs of the seven genes nanA, nanB, ply, lytA, lytB, lytC, and iga that have been associated to
bacterial survival in blood and immune evasion were variously present in the genomes [12, 16, 17,
24]. Both nanA and nanB gene homologs were identified in S. mitis RH50275_09 and S. mitis
RH50738_11; these were the only strains containing both neuraminidase genes. The nanA and nanB
homologs were neighbours. None of the S. mitis strains contained lytA and ply gene homologs
simultaneously. iga homologs were identified in all 14 S. oralis subsp. oralis whereas lytA
homologs only were identified in S. oralis subsp. oralis and subsp. tigurinus.

Polysaccharide capsule production (CPS) has been described important for bacterial avoidance of
the phagocytosis [19, 40]. Genes encoding homologs of Cps4 from S. pneumoniae TIGR4 were
identified in both S. mitis and S. oralis. cps4A gene homologs were present in all 40 strains whereas
genes homologous to cps4B, cps4C, and cps4D were variously present in the genomes. Eight S.
mitis strains and 22 S. oralis strains contained homologs of the four capsular genes cps4A, cps4B,
cps4C, and cps4D. Furthermore, 22 S. oralis strains and one S. mitis strain contained a gene
homologous to cps4I. One S. oralis subsp. dentisani strain, RH9883_08, contained genes
homologous to cps4E, cps4F, cps4J, cps4K, and cps4L.

In summary, three genes homologous to the adhesion genes, psaA, lmb and pavA were identified in
all 40 strains. The presence of the seven putative virulence genes (homologs of nanA, nanB, ply,
lytA, lytB, lytC and iga) important for immune evasion and colonisation in the 40 S. mitis and S.
oralis genomes were not coherent. A few S. oralis subspecies specific differences were observed. All 14 S. oralis subsp. oralis contained an iga homolog, whereas homologs of lytA only were identified in S. oralis subsp. oralis and S. oralis subsp. tigurinus. Homologs of nanB and ply were only identified in S. mitis. Furthermore, homologs to the cps4 genes were identified variously in S. oralis and S. mitis strains, but none of the strains included a full capsular locus compared to the VFDB reference S. pneumoniae TIGR4 genome.

Discussion

Assessment of virulence factors in clinical S. mitis and clinical S. oralis subsp. oralis, subsp. tigurinus and subsp. dentisani has only been sparsely conducted. In the present study, the functional profiles were extracted from 40 IE clinical strains of S. mitis and S. oralis subsp. oralis, subsp. tigurinus and subsp. dentisani, by using the pipeline PanFunPro [30]. We have previously used PanFunPro for extraction of a Mitis group streptococci core-genome for evaluation of core-genome phylogeny [27]. The core-genome phylogeny revealed a subclustering of S. oralis into three subclusters [27]. Subclustering of S. oralis was later illustrated by Jensen et al. [8] by using core-genome phylogeny and it was proposed that the species S. tigurinus and S. dentisani should be reassigned as subspecies in S. oralis. Core-genome phylogeny was basis for identification of the clinical IE strains in the present study and in addition, Fig. 1 clearly illustrates clustering of the S. oralis strains into the three subspecies.

The clustering of the three S. oralis subspecies strains in Fig. 1 based on the pan-genome indicates that other differences may occur between the subspecies than in the core-genes. By using a sequence identity percent > 40 % at protein level, few subspecies specific differences in virulence factors were observed between the three subspecies S. oralis subsp. oralis, subsp. tigurinus and
subsp. dentisani. The threshold at 40 % sequence identity was based on findings in a study by Rost [41] who described that 90 % of the protein pairs were homologous when using a cut-off at roughly 30% sequence identity. Furthermore, 40 % sequence identity has previously been used for protein identification in the Mitis group [42].

The alignment of the functional profiles against the VFDB database revealed that iga homologs were present in all 14 S. oralis subsp. oralis and in seven out of 12 S. mitis. The iga gene encoding IgA1 protease that cleaves the human immunoglobulin A1 in the hinge region, has been variously identified in S. mitis and S. oralis strains [8, 21, 42, 43]. IgA1 is a predominant immunoglobulin presented on the mucosal surfaces [44] and cleavage of this, limits the host humoral response and thereby promote colonisation of S. pneumoniae [12]. Recently, Jensen et al. [8] described that iga is only present in S. oralis subsp. oralis and not in S. oralis subsp. tigurinus and subsp. dentisani in accordance with the findings in the present study. These findings are further supported by Conrads et al. who used the former nomenclature and identified iga in S. oralis but not in S. tigurinus [45].

Another subspecies difference was observed between S. oralis subsp. oralis, subsp. tigurinus and subsp. dentisani in the present study (Table 1). Homologs of lytA were only identified in strains of S. oralis subsp. oralis and subsp. tigurinus. Conrads et al. did not include S. dentisani in their study [45]. lytA encodes the autolytic cell wall hydrolase Autolysin (LytA), which appears to be a predisposing circumstance for the release of cell cytoplasmic located protein pneumolysin (Ply) [46]. Pneumolysin (Ply) encoded by the gene ply, is a poreforming toxin that induces cell death by apoptosis. It is suggested to be an important factor for the initial establishment in nasal colonization and for development of septicemia [13, 14, 47]. The two genes lytA and ply have been localised simultaneously in all analysed S. pneumoniae genomes [24, 42] and in S. tigurinus AZ_3a [45]. In contrast, lytA and ply have only been identified in three out of 31 S. mitis genomes [24] and in none
of the examined *S. oralis* genomes [24, 42]. In the present study, only two *S. mitis* genomes contained genes homologous to *ply* and five genomes contained genes homologous to *lytA* (Table 1). *lytA* and *ply* homologs were not present simultaneously in any *S. mitis* strain, indicating that the presence and potential cooperation of *lytA* and *ply* is not a precondition for the *S. mitis* virulence.

Other cell wall hydrolases, (LytB and LytC), encoded by *lytB* and *lytC*, are important for the colonisation of *S. pneumoniae* in nasopharynx and they contribute to bacterial avoidance of phagocytosis mediated by neutrophils and alveolar macrophages [16, 48]. In the present study, *lytB* homologs were identified in all 28 *S. oralis* strains whereas genes homologous to *lytC* were identified in 14 of the *S. oralis* strains distributed on all three subspecies (Table 1). In contrast, genes homologous to both *lytB* and *lytC* were identified in the majority (11 out of 12) of the *S. mitis* strains. In strains where both genes were present, *lytB* and *lytC* homologs were located in different loci, indicating that these genes are not transcribed together.

Neuraminidase A and B (NanA and NanB) encoded by *nanA* and *nanB*, are other enzymes that have been stated important for colonisation and both enzymes seemed to be essential for survival in blood [17]. Intravenous infection with *nanA* and *nanB* mutants in mice, revealed a progressively clearance of bacteria in blood within 48 hours compared to the wild types, which persisted longer. In a previous study, *nanA* has been identified using PCR in all strains of *S. oralis* (*n* = 23) and *S. mitis* (*n* =10) [49], while only *nanB* was identified in strains of *S. mitis* by hybridization [25]. Genes homologous to *nanA* were identified in 27 strains of *S. oralis* and seven strains of *S. mitis* in the present study (Table 1). Genes homologous to *nanB* were only observed in six *S. mitis* strains in concordance with previous studies. Homologs of both *nanA* and *nanB* were only identified simultaneously in two *S. mitis* strains. In these strains *nanA* and *nanB* homologs were neighbours indicating that these two genes may belong to a *nanAB* locus which have been described in *S.*
pneumoniae [50]. Furthermore, the dispersed presence of nanA and nanB in S. mitis and S. oralis indicates that these two genes are not essential for the bacterial survival in blood.

Adhesion of bacterial cells to fibronectin may contribute to development of IE [51]. Fibronectin is an extracellular matrix protein secreted by a variety of cells and it is present in saliva and blood [52, 53]. S. pneumoniae adhere to immobilized fibronectin by the fibronectin binding surface protein PavA encoded by the gene pavA and it was demonstrated that pavA mutants had less ability to adhere to human epithelial and endothelial cells [18, 54]. A study of cell surface proteins in S. pneumoniae, S. mitis, and S. oralis showed that all 21 strains hybridized with pavA using microarray [55] and in another study pavA was identified in all S. tigurinus strains [45]. Lmb encoding the lipoprotein Lmb is another gene contributing to adhesion, described for Streptococcus agalactiae as a protein that mediates bacterial attachment to human laminin promoting transfer of bacteria to the bloodstream and colonisation of damaged epithelium [56]. The same study illustrated the presence of lmb in all 11 examined S. agalactiae serotypes, confirming the importance of this gene [56]. psaA encoding another lipoprotein PsaA also contributing to bacterial adhesion, was likewise identified in all serotypes of S. pneumoniae [20]. The virulence properties of psaA was described using in vitro studies where psaA- mutants illustrated significant less virulence compared to the wildtype when inoculated intranasal and intraperitoneal in mice [57]. As well S. pneumoniae as S. agalactiae strains have been associated with IE cases, though they are mostly associated with non-IE infections [11, 58].

In our study, genes homologues to pavA, lmb and psaA were identified in all 40 strains and these genes have been proven important for bacterial adhesion [54, 56, 59]. The presence of these genes across different species could be a result of horizontal gene transfer as earlier suggested by Zhang et al. for psaA [23].
Capsular polysaccharides (CPS) are indispensable for the virulence of *S. pneumoniae* by forming an inert shield, which prevent the phagocytosis [19, 40]. Today 97 serologically and structurally distinct CPS types have been recognised [60]. The encapsulated serotype 4 *S. pneumoniae* TIGR4 strain was used as reference in the present study to examine the presence of capsule loci in the 40 strains. The *cps* locus in TIGR4 include the genes *cps4A-cps4L* [61]. A *cps4A* homolog was identified in all 40 clinical strains (Table 1). Only one *S. oralis* subsp. *dentisani* strain (RH9883_08) contained genes homologous to *cps4E, cps4F, cps4J, cps4K,* and *cps4L*. Serotype switching between *S. mitis* strains and the *S. pneumoniae* TIGR4 strain has been reported before [62], which may also be possible for *S. oralis* subsp. *dentisani*. Skov *et al.* [63] identified complete *cps* loci in 74% of the 66 investigated *S. mitis* strains and in 95% of the 20 investigated *S. oralis* strains including the subspecies *tigurinus* and *dentisani*. They confirmed capsule expression using antigenic analyses and demonstrated serological identities with different pneumococcal serotypes [63]. In the present study, eight *S. mitis* strains and 22 *S. oralis* strains contained genes homologous to *cps4A, cps4B, cps4C,* and *cps4D*. The *cpsB-cpsD* have been found essential for encapsulation in *S. pneumoniae* whereas *cpsA* influenced the level of CPS produced [64]. The presence of *cps4A, cps4B, cps4C,* and *cps4D* homologs in the eight *S. mitis* and 22 *S. oralis* strains indicates that these strains might be able to express capsule proteins. However, identification of capsular genes is not synonymous with capsule expression. Similar antigenic analyses as conducted by Skov *et al.* [63] could elucidate whether the IE strains in the present study express capsules.

The former species *S. dentisani* now *S. oralis* subsp. *dentisani* were originally isolated from the oral cavity [65]. A recently study conducted by López-López *et al.* confirmed this by identifying *S. dentisani* in metagenomic sequences from 118 healthy individuals [6]. Beside the ability to colonize the oral cavity, the authors demonstrated that *S. dentisani* affects the growth of the oral pathogens *Streptococcus mutans, Streptococcus sobrinus* and *Prevotella intermedia,* illustrating a probiotic...
feature of *S. dentisani*. Based on their findings they proposed clinical trials to test the potential of *S. dentisani* in promoting human oral health [6]. In the present study, the isolation of six strains from IE patients, clearly demonstrates that *S. oralis* subsp. *dentisani* is an IE causing agent. This new knowledge is important as experimentally inoculation of *S. dentisani* into the oral cavity of healthy humans may affect their ability to develop IE.

**Conclusion**

In the present study, we describe for the first time that *S. oralis* subsp. *dentisani* is able to cause infective IE. The hierarchical clustering based on the pan-genome illustrates clustering of the *S. oralis* strains into subsp. *oralis*, subsp. *dentisani* and subsp. *tigurinus* indicating that other differences may occur between the subspecies than in the core-genomes.

Alignment of 40 clinical *S. oralis* (subsp. *oralis*, subsp. *dentisani* and subsp. *tigurinus*) and *S. mitis* genomes against the VFDB database revealed genes in the genomes homologous to virulence genes that contribute to bacterial avoidance of the immune system, colonisation and adhesion. Three genes homologous to *psaA*, *pavA* and *lmb* that contribute to adhesion were identified in all strains. The presence of adhesion genes in all strains indicates the importance of adhesion properties for *S. mitis* and *S. oralis*. Seven genes (homologs of *nanA*, *nanB*, *ply*, *lytA*, *lytB*, *lytC* and *iga*) contributing to colonisation and evasion of the immune system were variously identified in the strains. *iga* homologs were identified in *S. mitis* and all 14 *S. oralis* subsp. *oralis* whereas *lytA* homologs were identified in *S. mitis*, *S. oralis* subsp. *oralis* and *S. oralis* subsp. *tigurinus* indicating subspecies specific differences in *S. oralis* virulence. Genes homologous to the capsular genes *cps4* in *S. pneumoniae* TIGR4 were variously identified in the 40 strains. However, none of the strains contained a full *cps4* locus compared to *S. pneumoniae* TIGR4. The virulence gene profiles of the 40 clinical *S. mitis* and *S. oralis* (subsp. *oralis*, subsp. *dentisani* and subsp. *tigurinus*) contribute with important knowledge about the virulence of these species and new subspecies. However, a
further elucidation of expression studies and *in vivo* studies are necessary before the clinical relevance of the three new subspecies can be established.
Author statements

Funding
This work was supported by the Danish Heart Foundation (12-04-R90-A4024-22720 and 15-R99-A6040-22951), The A.P. Møller Foundation for the Advancement of Medical Science, The Foundation of Hans and Norah Buchard, The Foundation of Aase and Ejner Danielsen and The Region Zealand Foundation for Health Research.

Conflicts of interests
The authors declare that they have no conflicts of interest.

Ethical statement
Recognition of the streptococcal strains was as part of the routine diagnostic at Departments of Clinical Microbiology in Capital Region of Denmark, Region Zealand and Region of Southern Denmark. The strains were analysed anonymously in a retrospective manner and ethical approval and informed consent were thus, not required.


Table 1. Homologs of virulence genes in the 40 *S. oralis* and *S. mitis* strains.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Product</th>
<th><em>S. oralis</em> Identity</th>
<th><em>S. oralis</em> subsp. tigurinus</th>
<th><em>S. oralis</em> subsp. dentisani</th>
<th><em>S. mitis</em> Identity</th>
<th><em>S. mitis</em> subsp. oralis Identity</th>
<th><em>S. mitis</em> subsp. tigurinus Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>pavA</td>
<td>Adherence and virulence protein A</td>
<td>14/14</td>
<td>8/8</td>
<td>6/6</td>
<td>12/12</td>
<td>71-72</td>
<td>70-71</td>
</tr>
<tr>
<td>lmb</td>
<td>Laminin-binding surface protein</td>
<td>14/14</td>
<td>8/8</td>
<td>6/6</td>
<td>12/12</td>
<td>64-65</td>
<td>67-64</td>
</tr>
<tr>
<td>psaA</td>
<td>Pneumococcal surface adhesion A</td>
<td>14/14</td>
<td>8/8</td>
<td>6/6</td>
<td>12/12</td>
<td>92-94</td>
<td>94-97</td>
</tr>
<tr>
<td>nanA</td>
<td>Neuraminidase A</td>
<td>14/14</td>
<td>7/8</td>
<td>6/6</td>
<td>7/12</td>
<td>64-74</td>
<td>49-75</td>
</tr>
<tr>
<td>nanB</td>
<td>Neuraminidase B</td>
<td>0/14</td>
<td>0/8</td>
<td>0/6</td>
<td>6/12</td>
<td>51-98</td>
<td></td>
</tr>
<tr>
<td>ply</td>
<td>Pneumolysin</td>
<td>0/14</td>
<td>0/8</td>
<td>0/6</td>
<td>2/12</td>
<td>41-51</td>
<td></td>
</tr>
<tr>
<td>lytA</td>
<td>Autolysin</td>
<td>4/14</td>
<td>3/8</td>
<td>0/6</td>
<td>5/12</td>
<td>45-60</td>
<td>57-85</td>
</tr>
<tr>
<td>lytB</td>
<td>Cell Wall Hydrolase</td>
<td>14/14</td>
<td>8/8</td>
<td>6/6</td>
<td>11/12</td>
<td>47-55</td>
<td>45-69</td>
</tr>
<tr>
<td>lytC</td>
<td>Cell Wall Hydrolase</td>
<td>5/14</td>
<td>6/8</td>
<td>3/6</td>
<td>11/12</td>
<td>44-57</td>
<td>40-86</td>
</tr>
<tr>
<td>iga</td>
<td>IgA1 protease</td>
<td>14/14</td>
<td>0/8</td>
<td>0/6</td>
<td>7/12</td>
<td>42-52</td>
<td>40-74</td>
</tr>
</tbody>
</table>

*Number of strains in which the genes are present. ** Percentage of identical amino acids obtained using BLASTP.

Figure legends

**Fig. 1.** Hierarchical clustering of Pearson correlation coefficients determined from the presence/absence of functional profiles in the 40 strains. The heat map colour indicate the Pearson correlation coefficient between the strains; the darker colour, the higher correlation. The colour bars shows the individual species of the particular strain: *S. oralis* subsp. *oralis* (dark blue), *S. oralis* subsp. *tigurinus* (light blue), *S. oralis* subsp. *dentisani* (green) and *S. mitis* (red).