View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Online Research Database In Technology

DTU Library

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; Nielsen, Xiaohui C

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):
Rasmussen, L. H., Iversen, K. H., Dargis, R., Christensen, J. J., Skovgaard, O., Justesen, U. S., ... Nielsen, X. C. (2017). In silico assessment of virulence factors in strains of Streptococcus oralis and Streptococcus mitis isolated from patients with Infective Endocarditis. Journal of Medical Microbiology, 66, 1316-1323. DOI: 10.1099/jmm.0.000573

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.

1 **Title page**

2

3 **Title**

4 *In silico* assessment of virulence factors in strains of *Streptococcus mitis* and *Streptococcus oralis*

5 isolated from patients with Infective Endocarditis.

6

7 **Authors:**

8 Louise H. Rasmussen[1,2*], Katrine Højholt[1,7*], Rimtas Dargis[1], Jens Jørgen Christensen[1,8], Ole

9 Skovgaard[2], Ulrik S. Justesen[3,5], Flemming S. Rosenvinge[4], Claus Moser[5], Oksana Lukjancenko[6],

10 Simon Rasmussen[7] & Xiaohui C. Nielsen[1]

11

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

13

14 Louise Hesselbjerg Rasmussen[1,2] lohra@regionsjaelland.dk

15 Katrine Højholt[1,7] katrine@cbs.dtu.dk

16 Rimtas Dargis[1] rida@regionsjaelland.dk

17 Jens Jørgen Christensen[1,8] jejc@regionsjaelland.dk

18 Ole Skovgaard[2] olesk@ruc.dk

19 Ulrik Stenz Justesen[3,5] ujustesen@health.sdu.dk

20 Flemming Schønning Rosenvinge[4] flemming.rosenvinge@rsyd.dk

21 Claus Moser[5] moser@dadlnet.dk

22 Oksana Lukjancenko[6] oklu@food.dtu.dk

23 Simon Rasmussen[7] simon@cbs.dtu.dk

24 **Corresponding author:** Xiaohui Chen Nielsen[1] xcn@regionssjaelland.dk

25 [1]

26 **Affiliations and addresses:**

---

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

27 [1]Department of Clinical Microbiology, Slagelse Hospital, Ingemannsvej 46, 4200 Slagelse,
28 Denmark.
29 [2]Department of Science and Environment, Roskilde University, Universitetsvej 1, 4000 Roskilde,
30 Denmark.
31 [3]Department of Clinical Microbiology, Odense University Hospital, J.B. Winsløws Vej 21, 2, 5000
32 Odense C, Denmark.
33 [4]Department of Clinical Microbiology, Vejle Hospital, Kabbeltoft 25, 7100 Vejle, Denmark.
34 [5]Department of Clinical Microbiology, Rigshospitalet, University Hospital of Copenhagen,
35 Blegdamsvej 9, 2100 Copenhagen Ø, Denmark.
36 [6]National Food Institute, Technical University of Denmark, Søltofts plads, Building 221, Kgs
37 Lyngby, Denmark.
38 [7]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of
39 Denmark, Kemitorvet, Building 208, 2800 Kgs Lyngby, Denmark.
40 [8]Department of Clinical Medicine, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen
41 N, Denmark.
42
43

44 ## Abstract

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

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

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

61 **Conclusion.** Differences in presence of virulence factors between the three *S. oralis* subspecies
62 were observed. The virulence gene profiles of the 40 *S. mitis* and *S. oralis* (subsp. *oralis,* subsp.
63 *dentisani* and subsp. *tigurinus*) contribute with important knowledge of these species and new
64 subspecies.

65

66

67 **Keywords:** Mitis group streptococci - Comparative genomics - Virulence factors - Infective
68 Endocarditis - *Streptococcus mitis* - *Streptococcus oralis.*

69

70

71 # Introduction

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

82 *Streptococcus pneumoniae,* another member of the Mitis group, is the closest relative to *S. oralis*
83 and *S. mitis.* Besides colonising the human nasopharynx, *S. pneumoniae* also causes local infections
84 and serious life-threatening diseases, such as septicaemia, meningitis, pneumonia and more rare IE
85 [9-11]. Virulence genes contributing to colonisation (e.g. *nanA, nanB, lytA, lytB, lytC,* and *ply*),
86 contributing to evasion of the immune system (e.g. *iga, cps*) and contributing to adhesion (e.g. *psaA*
87 and *pavA*) have been discovered in *S. pneumoniae* [12-20]. In addition, many of these genes have
88 been identified in *S. mitis* and *S. oralis.*

89 The Immunoglobulin A1 (IgA1) protease has been observed in both *S. oralis* and *S. mitis,* though
90 variously present in both species [8, 21]. The gene encoding the pneumococcal surface adhesion A
91 (*psaA*) has been identified in all investigated *S. mitis* and *S. oralis* [22, 23] and horizontal *psaA* gene
92 transfer has been suggested among the species in the Mitis group [23]. The genes *ply* and *lytA* have
93 both been recognized in the genomes of a minority of *S. mitis* genomes, but not in the genomes of *S.*
94 *oralis* [24, 25]. In contrast, both *S. mitis* and *S. oralis* exhibit neuraminidase activity when grown in
95 Brain Heart Infusion broth [26]. A widespread presence of the gene *pavA* was observed in a study

96 where all nine included *S. mitis* and 11 *S. oralis* strains hybridized with *pavA* illustrating the
97 importance of adherence and virulence protein A (PavA) for oral streptococci [25].

98

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

105

106 Materials and methods

107 Bacterial strains

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

119

120 **Genome annotation**

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

129

130 **Hierarchical clustering of species**

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

137

138 **Prediction of putative virulence genes**

139 Basic Local Alignment Search Tool (BLASTP) [36] was applied to search the translated protein
140 sequences against Virulence Factors of Pathogenic Bacteria database (VFDB), (Accessed 25 August
141 2015) which contains various virulence factors from other streptococci, *Staphylococcus aureus* and
142 *Enterococcus faecalis* [37-39]. The threshold for hits were an e-value < 0.001, a bit score > 50 and a
143 sequence identity percent > 40 %. The best hit was based on highest bit score.

144

145 # Results

146 **Whole genome sequence characterisation**

147 The number of scaffolds from the *de novo* assembly ranged from 17-85 (*S. mitis)*, 20-41 (*S. oralis*
148 subsp. *dentisani*), 7-47 (*S. oralis* subsp. *oralis*) and 7-47 (*S. oralis* subsp. *tigurinus*). The estimated
149 sizes of the *S. mitis* and *S. oralis* genomes ranged from 1.8 Mb-2.1 Mb. Each functional profile was
150 considered based on a coding sequence. Between 1,692-2,083 functional profiles were predicted in
151 the 12 *S. mitis* strains and 1,734-2,035 functional profiles were predicted in the 28 *S. oralis* strains.
152 There was no subspecies specific differences between the number of functional profiles in the 28 *S.*
153 *oralis* strains. The GC content was slightly higher in *S. oralis* (40.75-41.50 %) than in *S. mitis*
154 (39.71-40.28 %). Number of scaffolds, N50, the longest sequences and the number of functional
155 profiles in the 40 *S. mitis* and *S. oralis* genomes are presented in Appendix A.
156
157 When clustering the strains based on presence/absence of the functional profiles, a tight cluster
158 containing the *S. mitis* were identified (Fig. 1). The *S. oralis* strains clustered into three subclusters,
159 which were congruent with earlier observed subclusters based on core-gene phylogeny [27].
160 Furthermore, the subclustering of *S. oralis* were congruent with the division of the strains into the
161 three subspecies *S. oralis* subsp. *oralis,* subsp. *tigurinus* and subsp. *dentisani* [8].
162 Two *S. oralis* strains (*S. oralis* B007274_11 and *S. oralis* Y11577_11) with high correlation were
163 isolated from the same patient within a day and should be considered as the same strain.
164

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

167
168 In order to determine the presence of virulence genes in *S. mitis* and *S. oralis* subsp. *oralis,* subsp.
169 *tigurinus* and subsp. *dentisani,* the functional profiles based on coding sequences in the 40 strains

170 were aligned against the VFDB database. The number of strains that contained the putative
171 virulence genes and the protein sequence identity to the VFDB reference sequence are specified in
172 Table 1. Genes encoding proteins homologous to Adherence and virulence protein A (PavA)
173 Laminin binding protein (Lmb) and Pneumococcal surface adhesion A (PsaA) were identified in all
174 40 strains.

175 Homologs of the seven genes *nanA, nanB, ply, lytA, lytB, lytC,* and *iga* that have been associated to
176 bacterial survival in blood and immune evasion were variously present in the genomes [12, 16, 17,
177 24]. Both *nanA* and *nanB* gene homologs were identified in *S. mitis* RH50275_09 and *S. mitis*
178 RH50738_11; these were the only strains containing both neuraminidase genes. The *nanA* and *nanB*
179 homologs were neighbours. None of the *S. mitis* strains contained *lytA* and *ply* gene homologs
180 simultaneously. *iga* homologs were identified in all 14 *S. oralis* subsp. *oralis* whereas *lytA*
181 homologs only were identified in *S. oralis* subsp. *oralis* and subsp. *tigurinus.*
182 Polysaccharide capsule production (CPS) has been described important for bacterial avoidance of
183 the phagocytosis [19, 40]. Genes encoding homologs of Cps4 from *S. pneumoniae* TIGR4 were
184 identified in both *S. mitis* and *S. oralis*. *cps4A* gene homologs were present in all 40 strains whereas
185 genes homologous to *cps4B, cps4C,* and *cps4D* were variously present in the genomes. Eight *S.*
186 *mitis* strains and 22 *S. oralis* strains contained homologs of the four capsular genes *cps4A, cps4B,*
187 *cps4C,* and *cps4D*. Furthermore, 22 *S. oralis* strains and one *S. mitis* strain contained a gene
188 homologous to *cps4I.* One *S. oralis* subsp. *dentisani* strain, RH9883_08, contained genes
189 homologous to *cps4E, cps4F, cps4J, cps4K,* and *cps4L.*

190

191 In summary, three genes homologous to the adhesion genes, *psaA, lmb* and *pavA* were identified in
192 all 40 strains. The presence of the seven putative virulence genes (homologs of *nanA, nanB, ply,*
193 *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

218 subsp. *dentisani*. The threshold at 40 % sequence identity was based on findings in a study by Rost
219 [41] who described that 90 % of the protein pairs were homologous when using a cut-off at roughly
220 30% sequence identity. Furthermore, 40 % sequence identity has previously been used for protein
221 identification in the Mitis group [42].

222

223 The alignment of the functional profiles against the VFDB database revealed that *iga* homologs
224 were present in all 14 *S. oralis* subsp. *oralis* and in seven out of 12 *S. mitis.* The *iga* gene encoding
225 IgA1protease that cleaves the human immunoglobulin A1 in the hinge region, has been variously
226 identified in *S. mitis* and *S. oralis* strains [8, 21, 42, 43]. IgA1 is a predominant immunoglobulin
227 presented on the mucosal surfaces [44] and cleavage of this, limits the host humoral response and
228 thereby promote colonisation of *S. pneumoniae* [12]. Recently, Jensen *et al.* [8] described that *iga* is
229 only present in *S. oralis* subsp. *oralis* and not in *S. oralis* subsp. *tigurinus* and subsp. *dentisani* in
230 accordance with the findings in the present study. These findings are further supported by Conrads
231 *et al.* who used the former nomenclature and identified *iga* in *S. oralis* but not in *S. tigurinus* [45].
232 Another subspecies difference was observed between *S. oralis* subsp. *oralis,* subsp. *tigurinus* and
233 subsp. *dentisani* in the present study (Table 1). Homologs of *lytA* were only identified in strains of
234 *S. oralis* subsp. *oralis* and subsp. *tigurinus.* Conrads *et al.* did not include *S. dentisani* in their study
235 but they identified *lytA* in some *S. oralis* and *S. tigurinus* strains, congruent with the present results
236 [45]. *lytA* encodes the autolytic cell wall hydrolase Autolysin (LytA), which appears to be a
237 predisposing circumstance for the release of cell cytoplasmic located protein pneumolysin (Ply)
238 [46]. Pneumolysin (Ply) encoded by the gene *ply,* is a poreforming toxin that induces cell death by
239 apoptosis. It is suggested to be an important factor for the initial establishment in nasal colonization
240 and for development of septicemia [13, 14, 47]. The two genes *lytA* and *ply* have been localised
241 simultaneously in all analysed *S. pneumoniae* genomes [24, 42] and in *S. tigurinus* AZ_3a [45] . In
242 contrast, *lytA* and *ply* have only been identified in three out of 31 *S. mitis* genomes [24] and in none

243 of the examined *S. oralis* genomes [24, 42]. In the present study, only two *S. mitis* genomes
244 contained genes homologous to *ply* and five genomes contained genes homologous to *lytA* (Table
245 1). *lytA* and *ply* homologs were not present simultaneously in any *S. mitis* strain, indicating that the
246 presence and potential cooperation of *lytA* and *ply* is not a precondition for the *S. mitis* virulence.

247

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

256

257 Neuraminidase A and B (NanA and NanB) encoded by *nanA* and *nanB,* are other enzymes that have
258 been stated important for colonisation and both enzymes seemed to be essential for survival in
259 blood [17]. Intravenous infection with *nanA* and *nanB* mutants in mice, revealed a progressively
260 clearance of bacteria in blood within 48 hours compared to the wild types, which persisted longer.
261 In a previous study, *nanA* has been identified using PCR in all strains of *S. oralis* (n = 23) and *S.*
262 *mitis* (n =10) [49], while only *nanB* was identified in strains of *S. mitis* by hybridization [25]. Genes
263 homologous to *nanA* were identified in 27 strains of *S. oralis* and seven strains of *S. mitis* in the
264 present study (Table 1). Genes homologous to *nanB* were only observed in six *S. mitis* strains in
265 concordance with previous studies. Homologs of both *nanA* and *nanB* were only identified
266 simultaneously in two *S. mitis* strains. In these strains *nanA* and *nanB* homologs were neighbours
267 indicating that these two genes may belong to a *nanAB* locus which have been described in *S.*

268 *pneumoniae* [50]. Furthermore, the dispersed presence of *nanA* and *nanB* in *S. mitis* and *S. oralis*
269 indicates that these two genes are not essential for the bacterial survival in blood.
270
271 Adhesion of bacterial cells to fibronectin may contribute to development of IE [51]. Fibronectin is
272 an extracellular matrix protein secreted by a variety of cells and it is present in saliva and blood [52,
273 53]. *S. pneumoniae* adhere to immobilized fibronectin by the fibronectin binding surface protein
274 PavA encoded by the gene *pavA* and it was demonstrated that *pavA* mutants had less ability to
275 adhere to human epithelial and endothelial cells [18, 54]. A study of cell surface proteins in *S.*
276 *pneumoniae*, *S. mitis*, and *S. oralis* showed that all 21 strains hybridized with *pavA* using
277 microarray [55] and in another study *pavA* was identified in all *S. tigurinus* strains [45]. *lmb*
278 encoding the lipoprotein Lmb is another gene contributing to adhesion, described for *Streptococcus*
279 *agalactiae* as a protein that mediates bacterial attachment to human laminin promoting transfer of
280 bacteria to the bloodstream and colonisation of damaged epithelium [56]. The same study illustrated
281 the presence of *lmb* in all 11 examined *S. agalactiae* serotypes, confirming the importance of this
282 gene [56]. *psaA* encoding another lipoprotein PsaA also contributing to bacterial adhesion, was
283 likewise identified in all serotypes of *S. pneumoniae* [20]. The virulence properties of *psaA* was
284 described using *in vitro* studies where *psaA*$^{-}$ mutants illustrated significant less virulence compared
285 to the wildtype when inoculated intranasal and intraperitoneal in mice [57]. As well *S. pneumoniae*
286 as *S. agalactiae* strains have been associated with IE cases, though they are mostly associated with
287 non-IE infections [11, 58].
288 In our study, genes homologues to *pavA*, *lmb* and *psaA* were identified in all 40 strains and these
289 genes have been proven important for bacterial adhesion [54, 56, 59]. The presence of these genes
290 across different species could be a result of horizontal gene transfer as earlier suggested by Zhang *et*
291 *al.* for *psaA* [23].
292

293 Capsular polysaccharides (CPS) are indispensable for the virulence of *S. pneumoniae* by forming an
294 inert shield, which prevent the phagocytosis [19, 40]. Today 97 serologically and structurally
295 distinct CPS types have been recognised [60]. The encapsulated serotype 4 *S. pneumoniae* TIGR4
296 strain was used as reference in the present study to examine the presence of capsule loci in the 40
297 strains. The *cps* locus in TIGR4 include the genes *cps4A-cps4L* [61]. A *cps4A* homolog was
298 identified in all 40 clinical strains (Table 1). Only one *S. oralis* subsp. *dentisani* strain (RH9883_08)
299 contained genes homologous to *cps4E, cps4F, cps4J, cps4K,* and *cps4L.* Serotype switching
300 between *S. mitis* strains and the *S. pneumoniae* TIGR4 strain has been reported before [62], which
301 may also be possible for *S. oralis* subsp. *dentisani.* Skov *et al.* [63] identified complete *cps* loci in
302 74 % of the 66 investigated *S. mitis* strains and in 95 % of the 20 investigated *S. oralis* strains
303 including the subspecies *tigurinus* and *dentisani*. They confirmed capsule expression using
304 antigenic analyses and demonstrated serological identities with different pneumococcal serotypes
305 [63]. In the present study, eight *S. mitis* strains and 22 *S. oralis* strains contained genes homologous
306 to *cps4A, cps4B, cps4C,* and *cps4D.* The *cpsB-cpsD* have been found essential for encapsulation in
307 *S. pneumoniae* whereas *cpsA* influenced the level of CPS produced [64]. The presence of *cps4A,*
308 *cps4B, cps4C,* and *cps4D* homologs in the eight *S. mitis* and 22 *S. oralis* strains indicates that these
309 strains might be able to express capsule proteins. However, identification of capsular genes is not
310 synonymous with capsule expression. Similar antigenic analyses as conducted by Skov *et al.* [63]
311 could elucidate whether the IE strains in the present study express capsules.
312
313 The former species *S. dentisani* now *S. oralis* subsp. *dentisani* were originally isolated from the oral
314 cavity [65]. A recently study conducted by López-López *et al.* confirmed this by identifying *S.*
315 *dentisani* in metagenomic sequences from 118 healthy individuals [6]. Beside the ability to colonize
316 the oral cavity, the authors demonstrated that *S. dentisani* affects the growth of the oral pathogens
317 *Streptococcus mutans, Streptococcus sobrinus* and *Prevotella intermedia,* illustrating a probiotic

318 feature of *S. dentisani.* Based on their findings they proposed clinical trials to test the potential of *S.*
319 *dentisani* in promoting human oral health [6]. In the present study, the isolation of six strains from
320 IE patients, clearly demonstrates that *S. oralis* subsp. *dentisani* is an IE causing agent. This new
321 knowledge is important as experimentally inoculation of *S. dentisani* into the oral cavity of healthy
322 humans may affect their ability to develop IE.
323

324 Conclusion

325 In the present study, we describe for the first time that *S. oralis* subsp. *dentisani* is able to cause
326 infective IE. The hierarchical clustering based on the pan-genome illustrates clustering of the *S.*
327 *oralis* strains into subsp. *oralis,* subsp. *dentisani* and subsp. *tigurinus* indicating that other
328 differences may occur between the subspecies than in the core-genes.
329 Alignment of 40 clinical *S. oralis* (subsp. *oralis,* subsp. *dentisani* and subsp. *tigurinus*) and *S. mitis*
330 genomes against the VFDB database revealed genes in the genomes homologous to virulence genes
331 that contribute to bacterial avoidance of the immune system, colonisation and adhesion. Three
332 genes homologous to *psaA, pavA* and *lmb* that contribute to adhesion were identified in all strains.
333 The presence of adhesion genes in all strains indicates the importance of adhesion properties for *S.*
334 *mitis* and *S. oralis.* Seven genes (homologs of *nanA, nanB, ply, lytA, lytB, lytC* and *iga*) contributing
335 to colonisation and evasion of the immune system were variously identified in the strains.
336 *iga* homologs were identified in *S. mitis* and all 14 *S. oralis* subsp. *oralis* whereas *lytA* homologs
337 were identified in *S. mitis, S. oralis* subsp. *oralis* and *S. oralis* subsp. *tigurinus* indicating subspecies
338 specific differences in *S. oralis* virulence. Genes homologous to the capsular genes *cps4* in *S.*
339 *pneumoniae* TIGR4 were variously identified in the 40 strains. However, none of the strains
340 contained a full *cps4* locus compared to *S. pneumoniae* TIGR4. The virulence gene profiles of the
341 40 clinical *S. mitis* and *S. oralis* (subsp. *oralis,* subsp. *dentisani* and subsp. *tigurinus*) contribute
342 with important knowledge about the virulence of these species and new subspecies. However, a

343 further elucidation of expression studies and *in vivo* studies are necessary before the clinical

344 relevance of the three new subspecies can be established.

345 ## Author statements

346 ### Funding

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

351

352 ### Conflicts of interests

353 The authors declare that they have no conflicts of interest.

354

355 ### Ethical statement

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

360 Reference List

361

362 1. **Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE**. Defining the normal bacterial flora of the oral
363 cavity. *J. Clin. Microbiol*. 2005;43:5721-5732.
364 2. **Smith DJ, Anderson JM, King WF, van HJ, Taubman MA**. Oral streptococcal colonization of infants.
365 *Oral Microbiol. Immunol*. 1993;8:1-4.
366 3. **Matsui N, Ito M, Kuramae H, Inukai T, Sakai A *et al.*** Infective endocarditis caused by multidrug-
367 resistant Streptococcus mitis in a combined immunocompromised patient: an autopsy case report.
368 *J. Infect. Chemother*. 2012.
369 4. **Renton BJ, Clague JE, Cooke RP**. Streptococcus oralis endocarditis presenting as infective discitis in
370 an edentulous patient. *Int. J. Cardiol*. 2009;137:e13-e14.
371 5. **Zbinden A, Aras F, Zbinden R, Mouttet F, Schmidlin PR *et al.*** Frequent detection of Streptococcus
372 tigurinus in the human oral microbial flora by a specific 16S rRNA gene real-time TaqMan PCR. *BMC*
373 *Microbiol*. 2014;14:231.
374 6. **Lopez-Lopez A, Camelo-Castillo A, Ferrer MD, Simon-Soro A, Mira A**. Health-Associated Niche
375 Inhabitants as Oral Probiotics: The Case of Streptococcus dentisani. *Front Microbiol*. 2017;8:379.
376 7. **Zbinden A, Mueller NJ, Tarr PE, Eich G, Schulthess B *et al.*** Streptococcus tigurinus, a novel member
377 of the Streptococcus mitis group, causes invasive infections. *J. Clin. Microbiol*. 2012;50:2969-2973.
378 8. **Jensen A, Scholz CF, Kilian M**. Re-evaluation of the taxonomy of the Mitis group of the genus
379 Streptococcus based on whole genome phylogenetic analyses, and proposed reclassification of
380 Streptococcus dentisani as Streptococcus oralis subsp. dentisani comb. nov., Streptococcus
381 tigurinus as Streptococcus oralis subsp. tigurinus comb. nov., and Streptococcus oligofermentans as
382 a later synonym of Streptococcus cristatus. *Int J Syst Evol Microbiol*. 2016;66:4803-4820.
383 9. **Cartwright K**. Pneumococcal disease in western Europe: burden of disease, antibiotic resistance
384 and management. *Eur J Pediatr*. 2002;161:188-195.
385 10. **Tuomanen EI, Austrian R, Masure HR**. Pathogenesis of pneumococcal infection. *N Engl J Med*.
386 1995;332:1280-1284.
387 11. **de Egea V, Munoz P, Valerio M, de Alarcon A, Lepe JA *et al.*** Characteristics and Outcome of
388 Streptococcus pneumoniae Endocarditis in the XXI Century: A Systematic Review of 111 Cases
389 (2000-2013). *Medicine (Baltimore)*. 2015;94:e1562.
390 12. **Janoff EN, Rubins JB, Fasching C, Charboneau D, Rahkola JT *et al.*** Pneumococcal IgA1 protease
391 subverts specific protection by human IgA1. *Mucosal Immunol*. 2014;7:249-256.
392 13. **Mitchell TJ, Dalziel CE**. The biology of pneumolysin. *Subcell Biochem*. 2014;80:145-160.
393 14. **Hotomi M, Yuasa J, Briles DE, Yamanaka N**. Pneumolysin plays a key role at the initial step of
394 establishing pneumococcal nasal colonization. *Folia Microbiol (Praha)*. 2016.
395 15. **Walker JA, Allen RL, Falmagne P, Johnson MK, Boulnois GJ**. Molecular cloning, characterization,
396 and complete nucleotide sequence of the gene for pneumolysin, the sulfhydryl-activated toxin of
397 Streptococcus pneumoniae. *Infect Immun*. 1987;55:1184-1189.
398 16. **Ramos-Sevillano E, Moscoso M, Garcia P, Garcia E, Yuste J**. Nasopharyngeal colonization and
399 invasive disease are enhanced by the cell wall hydrolases LytB and LytC of Streptococcus
400 pneumoniae. *PLoS One*. 2011;6:e23626.
401 17. **Manco S, Hernon F, Yesilkaya H, Paton JC, Andrew PW *et al.*** Pneumococcal neuraminidases A and
402 B both have essential roles during infection of the respiratory tract and sepsis. *Infect Immun*.
403 2006;74:4014-4020.
404 18. **Holmes AR, McNab R, Millsap KW, Rohde M, Hammerschmidt S *et al.*** The pavA gene of
405 Streptococcus pneumoniae encodes a fibronectin-binding protein that is essential for virulence.
406 *Mol Microbiol*. 2001;41:1395-1408.
407 19. **Hostetter MK**. Serotypic variations among virulent pneumococci in deposition and degradation of
408 covalently bound C3b: implications for phagocytosis and antibody production. *J Infect Dis*.
409 1986;153:682-693.

410 20. **Morrison KE, Lake D, Crook J, Carlone GM, Ades E *et al.*** Confirmation of psaA in all 90 serotypes of
411 Streptococcus pneumoniae by PCR and potential of this assay for identification and diagnosis. *J Clin*
412 *Microbiol*. 2000;38:434-437.
413 21. **Bek-Thomsen M, Poulsen K, Kilian M**. Occurrence and evolution of the paralogous zinc
414 metalloproteases IgA1 protease, ZmpB, ZmpC, and ZmpD in Streptococcus pneumoniae and related
415 commensal species. *MBio*. 2012;3.
416 22. **Jado I, Fenoll A, Casal J, Perez A**. Identification of the psaA gene, coding for pneumococcal surface
417 adhesin A, in viridans group streptococci other than Streptococcus pneumoniae. *Clin Diagn Lab*
418 *Immunol*. 2001;8:895-898.
419 23. **Zhang Q, Ma Q, Su D, Li Q, Yao W *et al.*** Identification of horizontal gene transfer and
420 recombination of PsaA gene in streptococcus mitis group. *Microbiol Immunol*. 2010;54:313-319.
421 24. **Morales M, Martin-Galiano AJ, Domenech M, Garcia E**. Insights into the Evolutionary Relationships
422 of LytA Autolysin and Ply Pneumolysin-Like Genes in Streptococcus pneumoniae and Related
423 Streptococci. *Genome Biol Evol*. 2015;7:2747-2761.
424 25. **Madhour A, Maurer P, Hakenbeck R**. Cell surface proteins in S. pneumoniae, S. mitis and S. oralis.
425 *Iran J Microbiol*. 2011;3:58-67.
426 26. **Kamio N, Imai K, Shimizu K, Cueno ME, Tamura M *et al.*** Neuraminidase-producing oral mitis group
427 streptococci potentially contribute to influenza viral infection and reduction in antiviral efficacy of
428 zanamivir. *Cell Mol Life Sci*. 2015;72:357-366.
429 27. **Rasmussen LH, Dargis R, Hojholt K, Christensen JJ, Skovgaard O *et al.*** Whole genome sequencing
430 as a tool for phylogenetic analysis of clinical strains of Mitis group streptococci. *Eur J Clin Microbiol*
431 *Infect Dis*. 2016.
432 28. **Li JS, Sexton DJ, Mick N, Nettles R, Fowler VG, Jr. *et al.*** Proposed modifications to the Duke criteria
433 for the diagnosis of infective endocarditis. *Clin Infect Dis*. 2000;30:633-638.
434 29. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M *et al.*** SPAdes: a new genome assembly
435 algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455-477.
436 30. **Lukjancenko O, Thomsen MC, Voldby Larsen M, Ussery DP**. PanFunPro: PAN-genome analysis
437 based on FUNctional PROfiles. *F1000Research*. 2013;2:1-19.
438 31. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW *et al.*** Prodigal: prokaryotic gene recognition
439 and translation initiation site identification. *BMC Bioinformatics*. 2010;11:119.
440 32. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY *et al.*** Pfam: the protein families database.
441 *Nucleic Acids Res*. 2014;42:D222-230.
442 33. **Haft DH, Selengut JD, White O**. The TIGRFAMs database of protein families. *Nucleic Acids Res*.
443 2003;31:371-373.
444 34. **Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C *et al.*** SUPERFAMILY--sophisticated comparative
445 genomics, data mining, visualization and phylogeny. *Nucleic Acids Res*. 2009;37:D380-386.
446 35. **Zdobnov EM, Apweiler R**. InterProScan--an integration platform for the signature-recognition
447 methods in InterPro. *Bioinformatics*. 2001;17:847-848.
448 36. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J *et al.*** BLAST+: architecture and
449 applications. *BMC Bioinformatics*. 2009;10:421.
450 37. **Chen L, Yang J, Yu J, Yao Z, Sun L *et al.*** VFDB: a reference database for bacterial virulence factors.
451 *Nucleic Acids Res*. 2005;33:D325-D328.
452 38. **Chen L, Xiong Z, Sun L, Yang J, Jin Q**. VFDB 2012 update: toward the genetic diversity and molecular
453 evolution of bacterial virulence factors. *Nucleic Acids Res*. 2012;40:D641-645.
454 39. **Yang J, Chen L, Sun L, Yu J, Jin Q**. VFDB 2008 release: an enhanced web-based resource for
455 comparative pathogenomics. *Nucleic Acids Res*. 2008;36:D539-542.
456 40. **Abeyta M, Hardy GG, Yother J**. Genetic alteration of capsule type but not PspA type affects
457 accessibility of surface-bound complement and surface antigens of Streptococcus pneumoniae.
458 *Infect Immun*. 2003;71:218-225.
459 41. **Rost B**. Twilight zone of protein sequence alignments. *Protein Eng*. 1999;12:85-94.

460 42. **Kilian M, Poulsen K, Blomqvist T, Havarstein LS, Bek-Thomsen M *et al.*** Evolution of Streptococcus
461 pneumoniae and its close commensal relatives. *PLoS. One*. 2008;3:e2683.
462 43. **Reinholdt J, Tomana M, Mortensen SB, Kilian M**. Molecular aspects of immunoglobulin A1
463 degradation by oral streptococci. *Infect. Immun*. 1990;58:1186-1194.
464 44. **Kett K, Brandtzaeg P, Radl J, Haaijman JJ**. Different subclass distribution of IgA-producing cells in
465 human lymphoid organs and various secretory tissues. *J Immunol*. 1986;136:3631-3635.
466 45. **Conrads G, Barth S, Mockel M, Lenz L, van der Linden M *et al.*** Streptococcus tigurinus is frequent
467 among gtfR-negative Streptococcus oralis isolates and in the human oral cavity, but highly virulent
468 strains are uncommon. *J Oral Microbiol*. 2017;9:1307079.
469 46. **Lock RA, Hansman D, Paton JC**. Comparative efficacy of autolysin and pneumolysin as immunogens
470 protecting mice against infection by Streptococcus pneumoniae. *Microb Pathog*. 1992;12:137-143.
471 47. **Benton KA, Everson MP, Briles DE**. A pneumolysin-negative mutant of Streptococcus pneumoniae
472 causes chronic bacteremia rather than acute sepsis in mice. *Infect Immun*. 1995;63:448-455.
473 48. **Garcia P, Gonzalez MP, Garcia E, Lopez R, Garcia JL**. LytB, a novel pneumococcal murein hydrolase
474 essential for cell separation. *Mol Microbiol*. 1999;31:1275-1281.
475 49. **King SJ, Whatmore AM, Dowson CG**. NanA, a neuraminidase from Streptococcus pneumoniae,
476 shows high levels of sequence diversity, at least in part through recombination with Streptococcus
477 oralis. *J Bacteriol*. 2005;187:5376-5386.
478 50. **Gualdi L, Hayre JK, Gerlini A, Bidossi A, Colomba L *et al.*** Regulation of neuraminidase expression in
479 Streptococcus pneumoniae. *BMC Microbiol*. 2012;12:200.
480 51. **Moreillon P, Que YA, Bayer AS**. Pathogenesis of streptococcal and staphylococcal endocarditis.
481 *Infect Dis Clin North Am*. 2002;16:297-318.
482 52. **Babu JP, Dabbous MK**. Interaction of salivary fibronectin with oral streptococci. *J Dent Res*.
483 1986;65:1094-1100.
484 53. **Wang Y, Ni H**. Fibronectin maintains the balance between hemostasis and thrombosis. *Cell Mol Life*
485 *Sci*. 2016.
486 54. **Pracht D, Elm C, Gerber J, Bergmann S, Rohde M *et al.*** PavA of Streptococcus pneumoniae
487 modulates adherence, invasion, and meningeal inflammation. *Infect Immun*. 2005;73:2680-2689.
488 55. **Madhour A, Maurer P, Hakenbeck R**. Cell surface proteins in S. pneumoniae, S. mitis and S. oralis.
489 *Iran J. Microbiol*. 2011;3:58-67.
490 56. **Spellerberg B, Rozdzinski E, Martin S, Weber-Heynemann J, Schnitzler N *et al.*** Lmb, a protein with
491 similarities to the LraI adhesin family, mediates attachment of Streptococcus agalactiae to human
492 laminin. *Infect Immun*. 1999;67:871-878.
493 57. **Berry AM, Paton JC**. Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for
494 virulence of Streptococcus pneumoniae. *Infect Immun*. 1996;64:5255-5262.
495 58. **Abid L, Charfeddine S, Kammoun S**. Isolated Streptococcus agalactiae tricuspid endocarditis in
496 elderly patient without known predisposing factors: Case report and review of the literature. *J*
497 *Saudi Heart Assoc*. 2016;28:119-123.
498 59. **Romero-Steiner S, Pilishvili T, Sampson JS, Johnson SE, Stinson A *et al.*** Inhibition of pneumococcal
499 adherence to human nasopharyngeal epithelial cells by anti-PsaA antibodies. *Clin Diagn Lab*
500 *Immunol*. 2003;10:246-251.
501 60. **Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP *et al.*** Pneumococcal Capsules and Their
502 Types: Past, Present, and Future. *Clin Microbiol Rev*. 2015;28:871-899.
503 61. **Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD *et al.*** Complete genome sequence of a virulent
504 isolate of Streptococcus pneumoniae. *Science*. 2001;293:498-506.
505 62. **Rukke HV, Kalluru RS, Repnik U, Gerlini A, Jose RJ *et al.*** Protective role of the capsule and impact
506 of serotype 4 switching on Streptococcus mitis. *Infect Immun*. 2014;82:3790-3801.
507 63. **Skov Sorensen UB, Yao K, Yang Y, Tettelin H, Kilian M**. Capsular Polysaccharide Expression in
508 Commensal Streptococcus Species: Genetic and Antigenic Similarities to Streptococcus
509 pneumoniae. *MBio*. 2016;7.

510 64. **Morona JK, Paton JC, Miller DC, Morona R**. Tyrosine phosphorylation of CpsD negatively regulates
511 capsular polysaccharide biosynthesis in streptococcus pneumoniae. *Mol Microbiol*. 2000;35:1431-
512 1442.
513 65. **Camelo-Castillo A, Benitez-Paez A, Belda-Ferre P, Cabrera-Rubio R, Mira A**. Streptococcus
514 dentisani sp. nov., a novel member of the mitis group. *Int J Syst Evol Microbiol*. 2014;64:60-65.

515

516

517

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

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

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

520

521 # Figure legends

522 **Fig. 1.** Hierarchical clustering of Pearson correlation coefficients determined from the

523 presence/absence of functional profiles in the 40 strains. The heat map colour indicate the Pearson

524 correlation coefficient between the strains; the darker colour, the higher correlation. The colour bars

525 shows the individual species of the particular strain: *S. oralis* subsp. *oralis* (dark blue), *S. oralis*

526 subsp. *tigurinus* (light blue), *S. oralis* subsp. *dentisani* (green) and *S. mitis* (red).