Comparative genomics among members of the *Streptococcus bovis*/*Streptococcus equinus* complex

Papadimitriou K.1,2, Papandreou N.C.3, Hamodrakas S.J.2, Ferreira S.3, Supply P.3,4, Pot B.3 and Tsakalidou E.1

1Laboratory of Dairy Research, Department of Food Science and Technology, Agricultural University of Athens, Iera Odos 75, 118 55 Athens, Greece.
2Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 157 01, Greece.
3Genoscreen, Genomic Platform and R&D, Campus de l’Institut Pasteur, 1 rue du Professeur Calmette, 59000 Lille, France.
4Institut Pasteur de Lille, Center for Infection and Immunity of Lille (CIL), F-59019 Lille, France.

Correspondence to: k.papadimitriou@aua.gr

### Abstract

Background: Today, only one streptococcal species, i.e. *Streptococcus thermophilus* is recognized as food-grade. Interestingly, other streptococci like *Streptococcus mitis* and *Streptococcus infantarius* belonging to the *Streptococcus bovis*/*Streptococcus equinus* complex (SBEC) are also found in food matrices. However, these species are phylogenetically related to *Streptococcus galolyticus* and *Streptococcus pasteurianus* that have been linked to endocarditis, bacteremia and colon cancer.

Objectives: To compare the available genomes of the members of the SBEC in order to shed light on their evolution and phylogenetic relative and to assess whether their pathogenic potential.

Methods: Comparative genomics analyses including full chromosome and CDS alignments, whole genome phylogeny and evaluation of gene content (e.g. core genome, singlets, etc.) was performed with appropriate bioinformatics tools.

Conclusions: Despite the fact that the four species of the SBEC were found highly related based on whole genome phylogeny, there were two different patterns of evolution among them. *Streptococcus pasteurianus*, *S. macedonicus* and *S. infantarius* seem to have undergone a reductive evolution process that resulted in significantly diminished genome sizes and increased percentages of potential pseudogenes when compared to *S. galolyticus*. In addition, *S. macedonicus* and *S. infantarius* seem to have lost several gene families linked to the ability of *S. galolyticus* to survive in the gastrointestinal tract of herbivores and to its pathogenicity. Our findings indicate differences in the ecological niche and the pathogenic potential among the four species.

### Sequencing the genome of *S. macedonicus* ACA-DC 198

1. **1st step:** Shotgun pyrosequencing with 454 GS FLX Titanium (≈100 contigs)
2. **2nd step:** paired-end pyrosequencing with 454 GS FLX titanium (7 scaffolds)
3. **3rd step:** gap-closure and polishing with Illumina sequencing using the HiSeq 2000 (1 chromosome and 1 plasmid)
4. **4th step:** validation of the overall assembly (≈20X coverage) with an NGS optical map

### Annotating the genome of *S. macedonicus* ACA-DC 198

1. **1st step:** initial annotation was performed with the EcoRa and the RAST pipeline
2. **2nd step:** annotations were manually compiled in one using Kodon software
3. **3rd step:** final corrections and quality assessment was performed using Gnome2RMPF (including predictions for potential pseudogenes)

### Comparative genomics of *S. macedonicus* ACA-DC 198

- The complete genome sequence of *S. macedonicus* offered new opportunities to investigate the properties of the species at the genomic level.
- The inclusion of *S. macedonicus* and *S. infantarius* as subgroups of *S. galolyticus* has been previously suggested (Schlegel et al., Int J Syst Evol Microbiol. 2003).
- The blasting analysis using *S. galolyticus* has not been formally accepted due to low DNA-DNA hybridization relatedness values (≤70%) (Schlegel et al. Int J Syst Evol Microbiol. 2003)

### Additional characteristics of the genomes under investigation

- **S. macedonicus**, *S. pasteurianus* and *S. infantarius* genomes are shaped by selective pressures that favor extensive gene loss events and genome decay processes when compared to the *S. galolyticus* genome.
- This property (i.e. genome decay) has been linked to the adaptation of bacteria to rich in nutrients environments as in the case of *S. thermophilus* adaptation to the milk environment.

### Niche-specific and pathogenicity genes presence/absence

<table>
<thead>
<tr>
<th>Gene function</th>
<th>Presence/absence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>- accessory pilin (pil3)</strong></td>
<td>√√√√√</td>
</tr>
<tr>
<td><strong>- major pilin (pil3)</strong></td>
<td>√√√√√</td>
</tr>
<tr>
<td><strong>- sortase C (pil3)</strong></td>
<td>√√√√√</td>
</tr>
</tbody>
</table>

### Conclusions

1. *S. macedonicus* is most probably a separate species from *S. galolyticus*.
2. In silico analysis of *S. macedonicus* ACA-DC 198 suggests that:
   - the strain is at the process of adapting to a rich in nutrients environment
   - it shows a diminished capacity to live and survive in the gut tract of herbivores
   - it has a diminished pathogenic potential compared to *S. galolyticus*

### Bibliography


---

**Fig. 1.** Alignment of the first 7 scaffolds of *S. macedonicus* ACA-DC 198 obtained after consequencing against the optical map (A). Validation of the final hybrid assembly after closure and polishing with Illumina sequencing (B).

**Fig. 2.** Gene map of *S. macedonicus* ACA-DC 198

**Fig. 3.** Comparative chromosome (A) and CDS (B) mapping of *S. macedonicus* ACA-DC 198 against related species

**Partner alignments of the chromosomes at the nucleotide or the CDS level revealed the degrees of synteny between each pair** (Fig. 3). The identity score at the nucleotide level of *S. macedonicus* against *S. galolyticus* and *S. pasteurianus* was around 75% and 76%, respectively. Even more, the identity score at the CDS level dropped radically, reaching 58% in the case of *S. macedonicus* against *S. galolyticus* and 63% in the case of *S. macedonicus* against *S. pasteurianus*. These values are not useful to directly determine the actual taxonomy of the three species. It is in fact, that they are quite low and they coincide with the law (65-70%) identity value of intergenomic DNA-DNA hybridization experiments reported previously. Reinforcing the notion that *S. macedonicus* and *S. galolyticus* should remain separate species.

**Acknowledgments:** This work was cofinanced by the European Social Fund and the National resources EPEAEK and IFATH through the Thales project.

**Full chromosome alignments were performed using local collarin blocks (LCBs) among the three species. The analysis revealed a mosaic pattern of homology (Fig. 4). Importantly, a significant portion of the genomic information has been conserved, since the pseudogene and gene gain events as well as the gene decay within the CDS of each genome clearly differ as a result of new genome decay, gene gain events as lateral gene transfer (LGT) must have played a role during the evolution of the three species. In addition, chromosome nucleotide composition analysis resulted in the number of G+C content to be similar, indicating the same amount of gene decay between the three species.**