Complete genome sequence of the dairy isolate Streptococcus macedonicus ACA-DC 198

Papadimitriou K.1, Papandreu N.C.2, Ferreira S.3, Supply P.3,4, Pot B.4 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 (CILL), F-59019 Lille, France

Correspondence to: kpapadimitriou@aua.gr

Abstract
Within the Streptococcus genus, only Streptococcus thermophilus is considered to be non-pathogenic due to its adaptation to the milk environment. Streptococcus macedonicus is also an intriguing streptococcal species since its most frequent source of isolation in dairy is fermented foods, mainly of dairy origin. Sequencing of S. macedonicus ACA-DC 198 genome was performed using a combination of 454 GS-FLX pyrosequencing and HiSeq 2000 Illumina sequencing. The hybrid assembly between 454 and HiSeq2000 data (>100 contigs) resulted in a complete genomic scaffold of 2,328,674 bp and a finished of 1,170,654 bp. This genome assembly was validated against a Rhizobium optical map of the S. macedonicus genome. Sequences were annotated with the BabySeq and the RAST pipeline and manually curated using Kodon. Final corrections were made based on the quality assessment of the annotation using GenePRIMP. We found 2,113 protein-coding genes on the chromosome, 192 of which were identified as potential pseudogenes, indicating an ongoing genome decay process. This hypothesis is also supported by the approximately 220 kb smaller genome size of S. macedonicus compared to the S. pyogenes/gallolyticus, despite the high level of gene synteny between the two species. Such a reducing evolutionary process is common for lactic acid bacteria domesticated to the food environment, which in the case of S. thermophilus was also accompanied by the loss of pathogenicity traits. With our in silico analysis we attempt to investigate whether S. macedonicus shows traits that would support its adaptation to the dairy environment at the genomic level.

Sequencing the genome of S. macedonicus ACA-DC 198

• 1st step: shotgun pyrosequencing with 454 GS-FLX titanium (>100 contigs)
• 2nd step: 36 bp paired-end pyrosequencing with 454 GS-FLX titanium (7 scaffolds)
• 3rd step: gap-closure and polishing with Illumina sequencing using the HiSeq 2000 (1 chromosome and 1 plasmid)
• 4th step: validation of the overall assembly (>90% coverage) with an Rhizobium optical map

Annotating the genome of S. macedonicus ACA-DC 198

• 1st step: initial annotation was performed with the BabySeq and the RAST pipelines
• 2nd step: annotations were manually compiled in one using Kodon software
• 3rd step: final corrections and quality assessment was performed using GenePRIMP (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 scale.
The inclusion of S. macedonicus and S. pneumoniae as subspecies of S. pyogenes has been previously suggested (Schnittger et al., Int J Syst Evol Microbiol. 2003), but this lacticomycological status has not been formally accepted due to low DNA-DNA hybridization similarities (~72%) (Mikulski et al., Int J Syst Evol Microbiol. 2002).

Conclusions
Our findings clearly suggest that not only S. macedonicus, but also S. pneumoniae and S. infantarius have diverged from S. pyogenes in their potential to elaborate complex plant carbohydrate and to cope with the harsh environment of the GI tract of mammals. Further analysis of the S. macedonicus genome revealed that it contains key pili-implantation genes that are absent in P. gingivalis (pilA, pilB, pilC, pilD, pilF). Among these genes, pilA, pilB and pilC are essential for the adhesion and colonization of S. macedonicus to epithelial cells and also to oral tissues, as well as in the induction of IL-1β production by TH2 cells in the mouse model. Our findings also suggest that S. macedonicus may have a role in the development of dental plaque and the induction of immune responses in the host.

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

Fig. 2. Genome map of S. macedonicus ACA-DC 198

• 16S-rRNA gene
• 18S rRNA gene
• 23S rRNA gene
• 70S rRNA gene

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

Fig. 4. Complete alignment of the S. macedonicus ACA-DC 198 genome against those of related species using local outlier blocks (LOBs)

Fig. 5. Venn diagram of S. macedonicus ACA-198 genome and those of related species (A): UPGA tree of all currently available streptococcal genomes (B)

Additional characteristics of the genomes under investigation

Table 1. Characteristic genes presence/absence

Niche-specific and pathogenicity genes presence/absence

Table 2. Niche-specific and pathogenicity genes presence/absence

Bibliography