Comparative genomic analysis among three dairy *Streptococcus macedonicus* strains

Maria Kazou[1], Voula Alexandraki[1], Athanasia Sarafianou[1], Bruno Pot[2,3,4,5], Pierre Renault[6,7], Effie Tsakalidou[1], Konstantinos Papadimitriou[1]

[1]*Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Athens, Greece*
[2]*Institut Pasteur de Lille, Center for Infection and Immunity of Lille (CIIL), Lille, France*
[3]*Inserm U1019, Lille, France*
[4]*CNRS UMR8204, Lille, France*
[5]*Univ Lille de Nord France, Lille, France*
[6]*INRA, UMR1319 Micalis, Jouy-en-Josas F-78352, France*
[7]*AgroParisTech, UMR Micalis, Jouy-en-Josas F-78352, France*

Abstract

Streptococcus macedonicus and *Streptococcus infantarius*, two species within the *Streptococcus bovis*/*Streptococcus equinus* complex (SBSEC) are frequently found in spontaneously fermented foods especially of dairy origin. These two species have been suggested to be non-pathogenic and to have been adapted to the dairy environment similarly to *Streptococcus thermophilus*. Analysis of the first complete genomes of *S. macedonicus* and *S. infantarius* indicated that they may be indeed adapted to milk but they also have retained a restricted repertoire of virulence traits when compared to well characterized pathogenic streptococci. In this study we performed comparative genomic analysis among *S. macedonicus* strains, isolated from Italian, French and Greek dairy products. The contigs of the Italian and French partial genomes were aligned using as a reference the Greek *S. macedonicus* ACA-DC 198 genome, the only *S. macedonicus* genome that is completely sequenced to date. Based on these assemblies, we constructed two distinct pseudochromosomes for the French and the Italian strains. Despite the artifactual nature of the two chromosomes, pairwise alignments among the three genomes revealed a high degree of synteny. The genetic information was overall conserved, but strain specific regions also existed. Furthermore, the analysis revealed that the French *S. macedonicus* strain has lost genes involved in the catabolism of complex plant carbohydrates, in the adhesion to the host's cells and in haemolysis. On the other hand, an extra lactose operon and a proteolytic system characteristic of lactic acid bacteria were identified, indicating their evolutionary adaptation to the milk environment. Analogous observations could also be made for the Italian strain. Even though our findings further support the adaptation of the *S. macedonicus* species to milk fermentation, the analysis of the additional *S. macedonicus* genomes, including non-dairy isolates, may be necessary to clarify its pathogenic potential

Results and Discussion

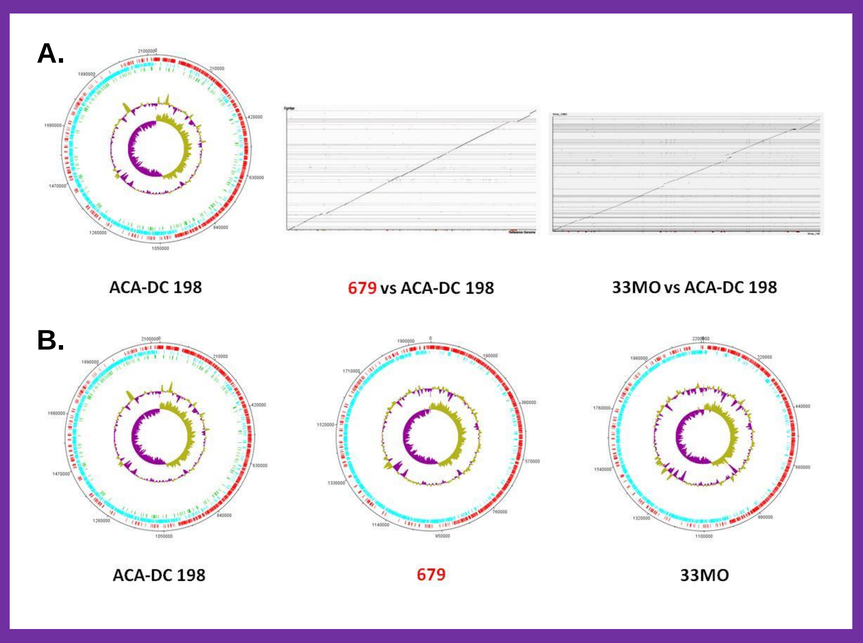


Figure 1 – A. The contigs of the two partial genomes were aligned using as a reference the complete genome of the *S. macedonicus* ACA-DC 198 strain. B. Based on the alignments, the pseudochromosomes and the relevant maps were constructed.

Figure 2 – Chromosome alignments of the *S. macedonicus* strains were calculated by progressiveMauve. The local collinear blocks (LCBs) of conserved sequences among the strains are represented by rectangles of the same colour. Connecting lines can be used to visualize synteny or rearrangement. LCBs positioned above or under the chromosome (black line) correspond to the forward and reverse orientation, respectively. The level of conservation is equivalent to the level of vertical colour filling within the LCBs (e.g. white regions are strain-specific). Sequences not placed within an LCB are unique for the particular strain.

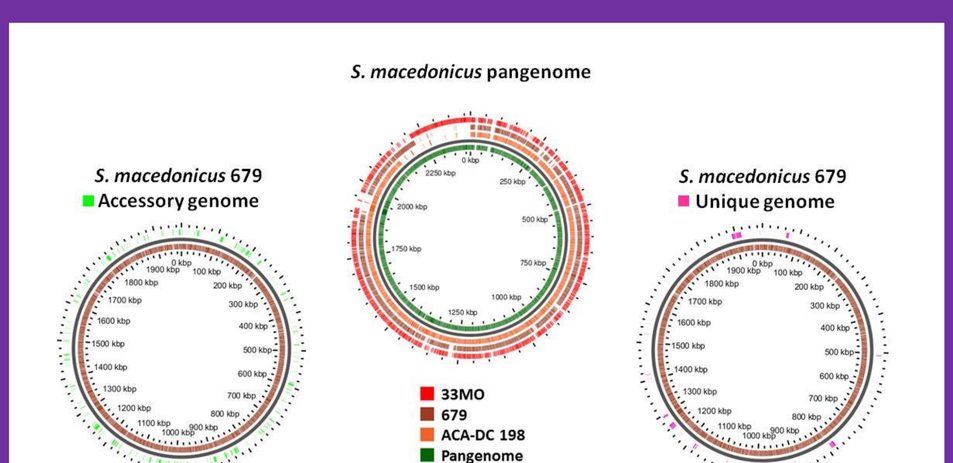


Figure 3 – Further analysis showed that the pangenome of the *S. macedonicus* strains is comprised of approximately 2400 genes. Concerning the genome of 679 strain, a great number of genes were distributed among the core and the accessory genome, while the unique genome contained 250 genes.

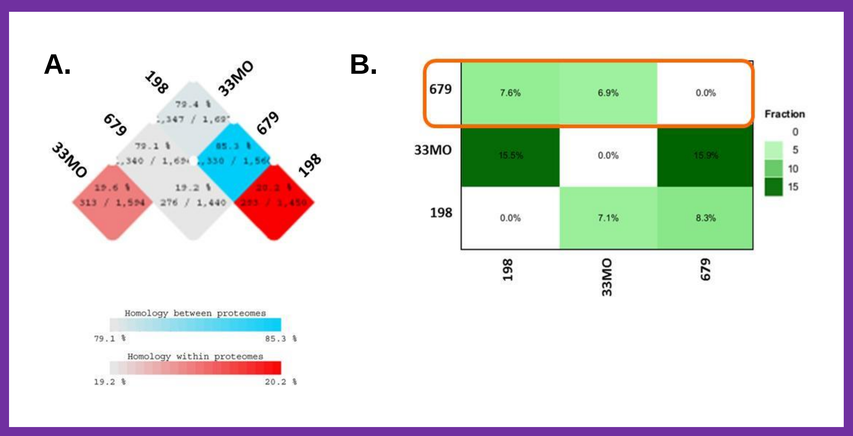


Figure 4 – A. Pairwise pan and core comparison between the three *S. macedonicus* strains, showed that the pairs share 80-85% of the protein families. Homology estimation within the proteome, revealed that approximately 19% of protein families in the proteome of the 679 strain has more than one member. B. Proteomes were also compared in pairs to estimate the fraction of specific proteins which are present in one genome and absent in another. The proteome of the 679 strain has 7% specific proteins.

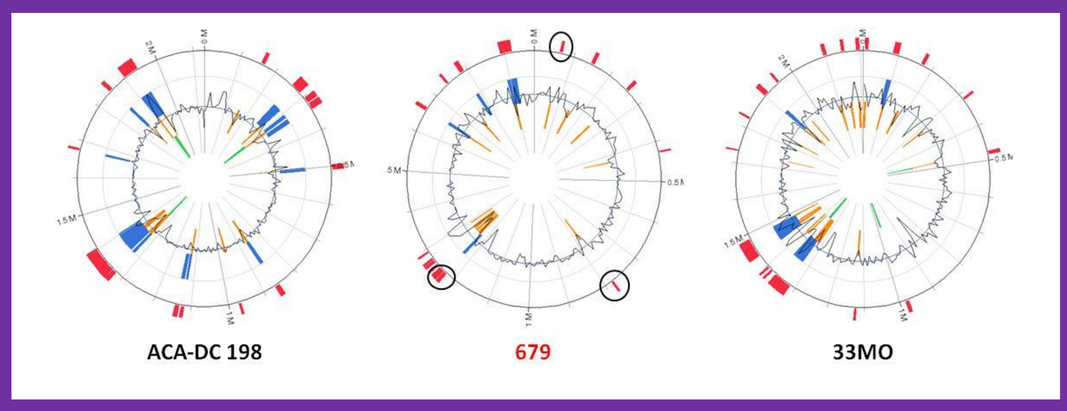
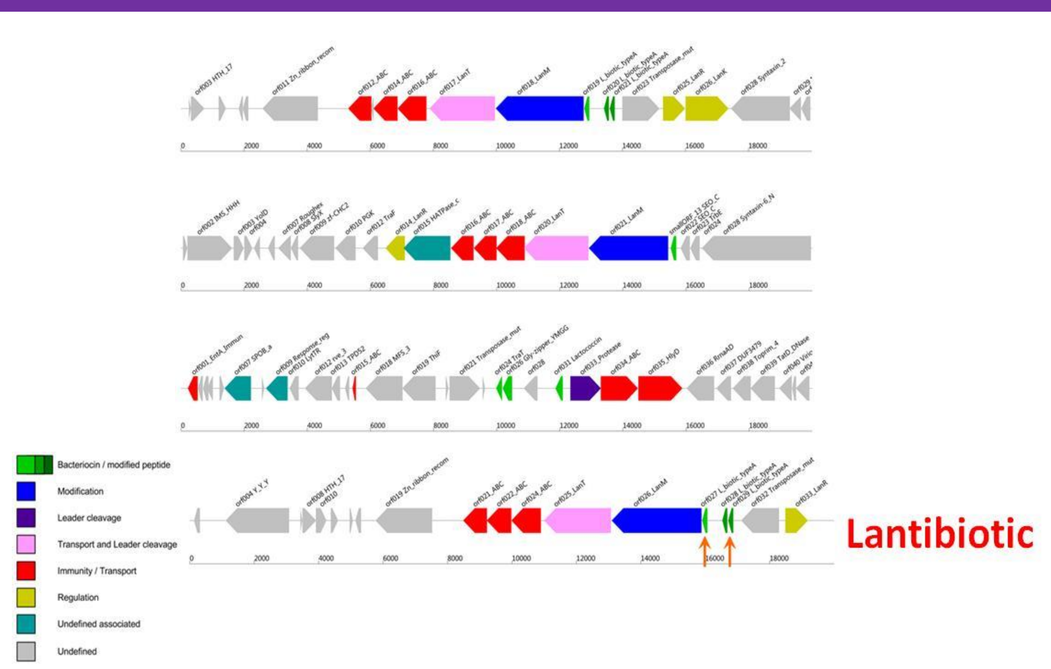


Figure 5 – A great number of genomic islands in the three *S. macedonicus* strains were found. Specifically for the 679 strain we identified 5 unique genomic island, as indicated by the black circles.



CRISPR arrays of *S. macedonicus* strains		
Strain	**CRISPR**	**Spacers**
679	1	11
ACA-DC 198	1	49
33MO	2	3
		36

Figure 6 – A potential lantibiotic were predicted and BLAST analysis of the entire gene cluster showed high identity to the known bacteriocin macedocin of the ACA-DC 198 strain. The CRISPR analysis revealed the existence of 1 CRISPR and low number of spacers

Conclusions

1. Our findings support two distinct evolutionary patterns within the SBSEC. On the one hand, *S. gallolyticus* is a species without apparent genome decay and the available genomes suggest that it is a robust bacterium able to thrive in the rumen of herbivores. On the other hand, the remaining SBSEC species, i.e. *S. macedonicus*, *S. pasteurianus* and *S. infantarius* exhibit decreased genome sizes accompanied by increased percentages of potential pseudogenes due to extensive genome decay, suggesting adaptation to nutrient-rich environments. This does not necessarily mean that the environment to which the three species have been adapted is the same.

2. It has been proposed that members of the SBSEC like *S. gallolyticus* may be part of the etiology of colon cancer by causing chronic inflammation. In order to assess the pathogenicity of this group of streptococci, more research is needed on the specific mechanisms employed by SBSEC members to cause disease. More comparative and functional genomics studies comprising SBSEC genomes are necessary that will cover additional species of the complex. New clinico-epidemiological studies should also be undertaken in view of the most recent changes in the taxonomy of the SBSEC complex.

Bibliography

- Papadimitriou K, Anastasiou R, Mavrogonatou E, Blom J, Papandreou NC, Hamodrakas SJ, Ferreira S, Renault P, Supply P, Pot B, Tsakalidou E., "Comparative genomics of the dairy isolate *Streptococcus macedonicus* ACA-DC 198 against related members of the *Streptococcus bovis*/*Streptococcus equinus* complex," BMC Genomics, 2014, vol. 15, pp. 272.
- Papadimitriou K, Ferreira S, Papandreou NC, Mavrogonatou E, Supply P, Pot B, Tsakalidou E., "Complete genome sequence of the dairy isolate *Streptococcus macedonicus* ACA-DC 198," J Bacteriology, 2012, vol. 194, pp. 1838-9.
- Papadimitriou K, Anastasiou R, Maistrou E, Plakas T, Papandreou NC, Hamodrakas SJ, Ferreira S, Supply P, Renault P, Pot B, Tsakalidou E., "Acquisition of plasmid pSMA198 by *Streptococcus macedonicus* ACA-DC 198 through horizontal gene transfer supports the dairy origin of the strain," PLOS One, under review.

Acknowledgments

This work was financially supported by the THALES Program (Funded by the Greek General Secretariat for Research and Technology).

Γεωπονικό Πανεπιστήμιο Αθηνών
Agricultural University of Athens

NSRF 2007-2013 programme for development