Comparative genomics among dairy strains of *Streptococcus thermophilus*

Voula Alexandraki[1], Maria Kazou[1], Nikos C. Papandreou[2], Stavros J. Hamodrakas[2], Bruno Pot[3, 4, 5, 6], Effie Tsakalidou[1] and Konstantinos Papadimitriou[1,*]

[1] Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens,Iera Odos 75, 118 55 Athens, Greece
[2] Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, 157 01 Athens, Greece
[3] Institut Pasteur de Lille, Center for Infection and Immunity of Lille (CIIL), F-59019 Lille, France
[4] Inserm U1019, F-59019 Lille, France
[5] CNRS UMR8204, F-59021 Lille, France
[6] Univ Lille de Nord France, F-59000 Lille, France

**Correspondance to: kpapadimitriou@aua.gr*

Abstract

Microorganisms like lactic acid bacteria are employed for the biotransformation of raw materials into fermented foods. Fermented foods have increased nutritional value and shelf-life as well as improved organoleptic characteristics compared to the raw materials. Interestingly, there are several genera within lactic acid bacteria that are considered to be important for food fermentations including the *Streptococcus* genus. However, only *Streptococcus thermophilus* is used as a starter culture. *Streptococcus thermophilus* has been adapted to milk and dairy products through a reductive evolution process that has led to the loss of typical streptococcal pathogenic traits. In this work we present the comparative genomic analysis among the recently sequenced genome of *S. thermophilus* ACA-DC 29 isolated from yogurt and the existing seven complete genome sequences of *S. thermophilus*. Full chromosome alignments revealed a high degree of synteny among the different strains although strain specific differences could also be observed. The pangenome of the eight strains was comprised of approximately 2,300 genes. Concerning the ACA-DC 29 strain, the majority of genes was distributed in the core and the accessory genomes. We also identified a significant percentage of unique genes, i.e. approximately 250, involved in various biological processes. Further analysis of these unique genes revealed that several of them may have been acquired through horizontal gene transfer. We also predicted five potential antimicrobial peptides and two CRISPR systems, which may confer resistance against phages. Overall, our analysis provides useful insights into the technological potential of the ACA-DC 29 strain.

Results and Discussion

Figure 1 - The circular map of the genome of *Streptococcus thermophilus* ACA-DC 29. Genomic features appearing from the periphery to the centre of the map: 1. Forward CDSs (red); 2. Reverse CDSs (cyan); 3. %GC plot; 4. GC skew.

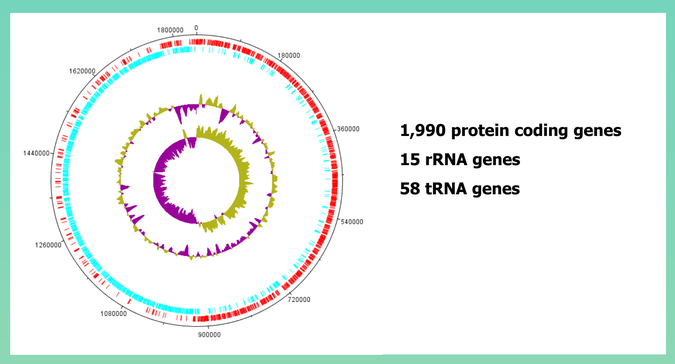


Figure 2 – A. Pairwise pan and core comparison between *Streptococcus thermophilus* ACA-DC 29 genome sequence and the seven additional complete genome sequences from dairy *Streptococcus thermophilus*. Pairs of genomes share 80% of the protein families. Homology estimation within the proteome of *Streptococcus thermophilus* ACA-DC 29 revealed that approximately 17% of the protein families had more than one member

B. Pairwise proteome comparison between *Streptococcus thermophilus* ACA-DC 29 genome sequence and the seven additional complete genome sequences from dairy *Streptococcus thermophilus*. On average the proteome of *Streptococcus thermophilus* ACA-DC 29 contained 12-15% specific proteins

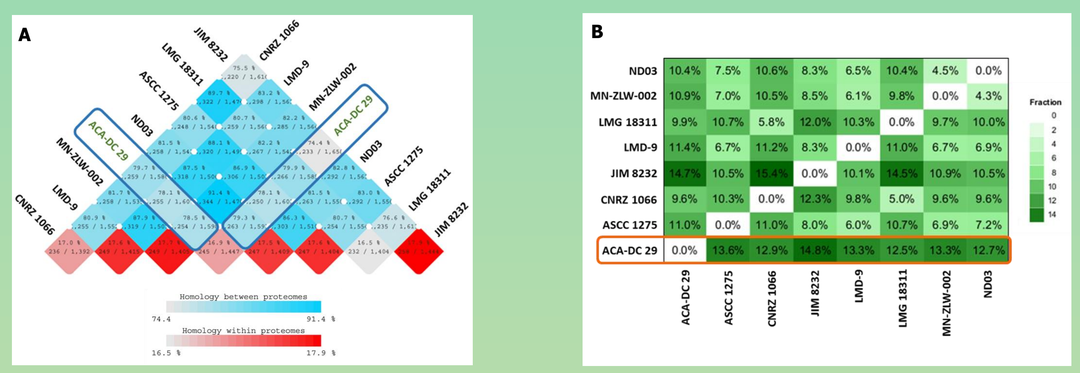


Figure 3 – Chromosome alignments of the *Streptococcus thermophilus* strains as calculated by progressiveMauve. Local collinear blocks (LCBs) of conserved sequences among the strains are represented by rectangles of the same colour. Connecting lines can be used to visualize synteny or rearrangement. 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.

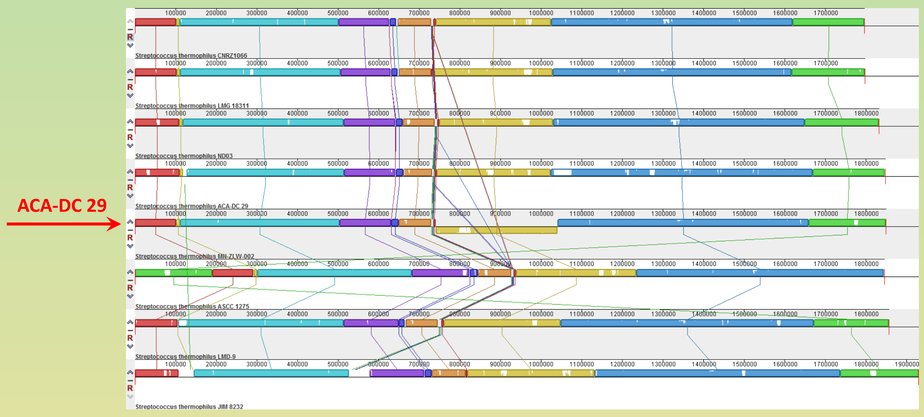


Figure 4 – Circular map of the *Streptococcus thermophilus* 29 genome. Highlighted regions correspond to genomic islands (GIs). GIs are colored within the circular maps according to the tool that predicted each one: green, orange and blue were predicted with IslandPick, SIGI-HMM and IslandPath-DIMOB, respectively. The integrated GIs are presented on the periphery in red. The black line plot represents the GC content (%) of the genomic sequences. Numbering of the GIs for each genome starts from the first GI found after position 0 of the genome and going clockwise. As seen below, from the centre to the periphery of the map, 3 GIs were predicted using IslandPick, 8 GIs using SIGI-HMM and 4 GIs using the IslandPath-DIMOB tool. Finally, 12 GIs are integrated.

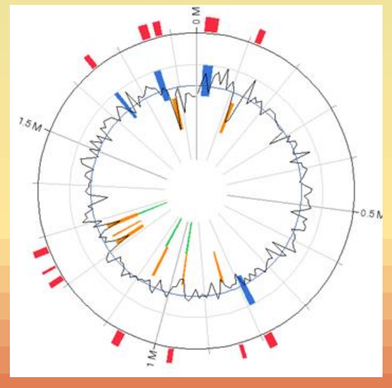


Figure 5 – A. The pangenome of the eigth *Streptococcus thermophilus* strains is comprised of approximately 2,300 genes. B. The core genome, the accessory genome and the unique genes of *Streptococcus thermophilus* ACA-DC 29 strain

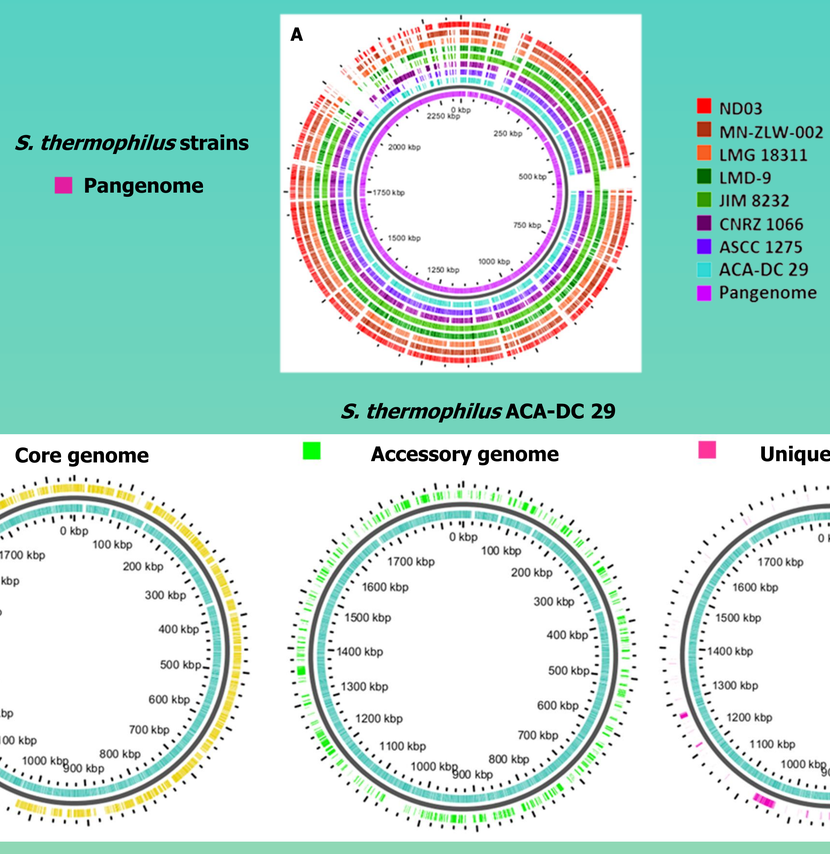


Figure 6 – A. Biological process distribution of the 250 unique genes of *S. thermophilus* ACA-DC 29. B. Top hits species distribution of the 250 unique genes of *S. thermophilus* ACA-DC 29

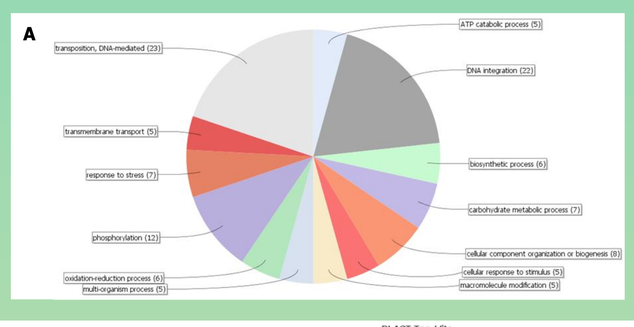
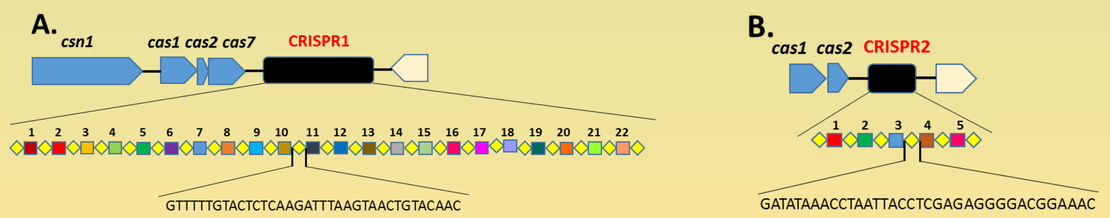


Figure 7 – Genes encoding for antimicrobial peptides in the genome sequence of *S. thermophilus* ACA-DC 29 as predicted by BAGEL3

Figure 8 – CRISPR systems in the genome sequence of *S. thermophilus* ACA-DC 29 as predicted by CRISPRfinder



Bibliography

• **Alexandraki V, Kazou M, Papandreou NC, Hamodrakas SJ, Pot B, Tsakalidou E, Papadimitriou K, "Complete genome sequence of the dairy isolate *Streptococcus thermophilus* ACA-DC 29," in preparation.**

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