

Characterization of the dairy *Streptococcus thermophilus* ACA-DC 29 strain through comparative genomics

Voula Alexandraki¹, Maria Kazou¹, Pierre Renault², Bruno Pot³, Effie Tsakalidou¹ and Konstantinos Papadimitriou^{1,*}

¹ Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Athens, Greece

² INRA, UMR1319 Micalis, Jouy-en-Josas F-78352, France

³ Institut Pasteur de Lille, Center for Infection and Immunity of Lille (CILL), Lille, France

*e-mail: kpapadimitriou@aua.gr

Abstract

Background: Although the *Streptococcus* genus includes mainly pathogenic species, *Streptococcus thermophilus* is a widely used dairy starter culture of great economic importance for the food industry. *S. thermophilus* has been adapted to milk probably through a degenerative evolution process that has led to the loss of typical streptococcal pathogenic traits.

Objectives: The genome sequence of the yogurt isolate *S. thermophilus* ACA-DC 29 was analyzed for assessing the technological potential of this strain. Comparative genomics analysis was also performed between the genome of ACA-DC 29 and the existing complete genome sequences of *S. thermophilus*.

Methods: The genome sequence of ACA-DC 29 was annotated using online annotation tools. Full chromosome alignments were calculated with Progressive Mauve. The pangenome, the core genome and the unique genes were predicted with the GView Server. The genomic islands, the CRISPRs and the antimicrobial peptides were predicted with IslandViewer, CRISPRcomp and BAGEL3, respectively.

Conclusions: The analysis of the *S. thermophilus* ACA-DC 29 genome sequence revealed the absence of pathogenic features. Genes related to the adaptation to milk were identified. Full chromosome alignments showed a high degree of synteny among the different strains. The pangenome of the ten strains comprised of approximately 2,300 genes. Concerning the ACA-DC 29 strain, approximately 250 unique genes involved in various biological processes were also identified. Further analysis indicated that several of them may have been acquired through horizontal gene transfer. Five potential antimicrobial peptides and two CRISPR systems, which may confer resistance against phages, were also predicted.

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 center of the map: 1. Forward CDSs (red); 2. Reverse CDSs (cyan); 3. %GC plot; 4. GC skew.

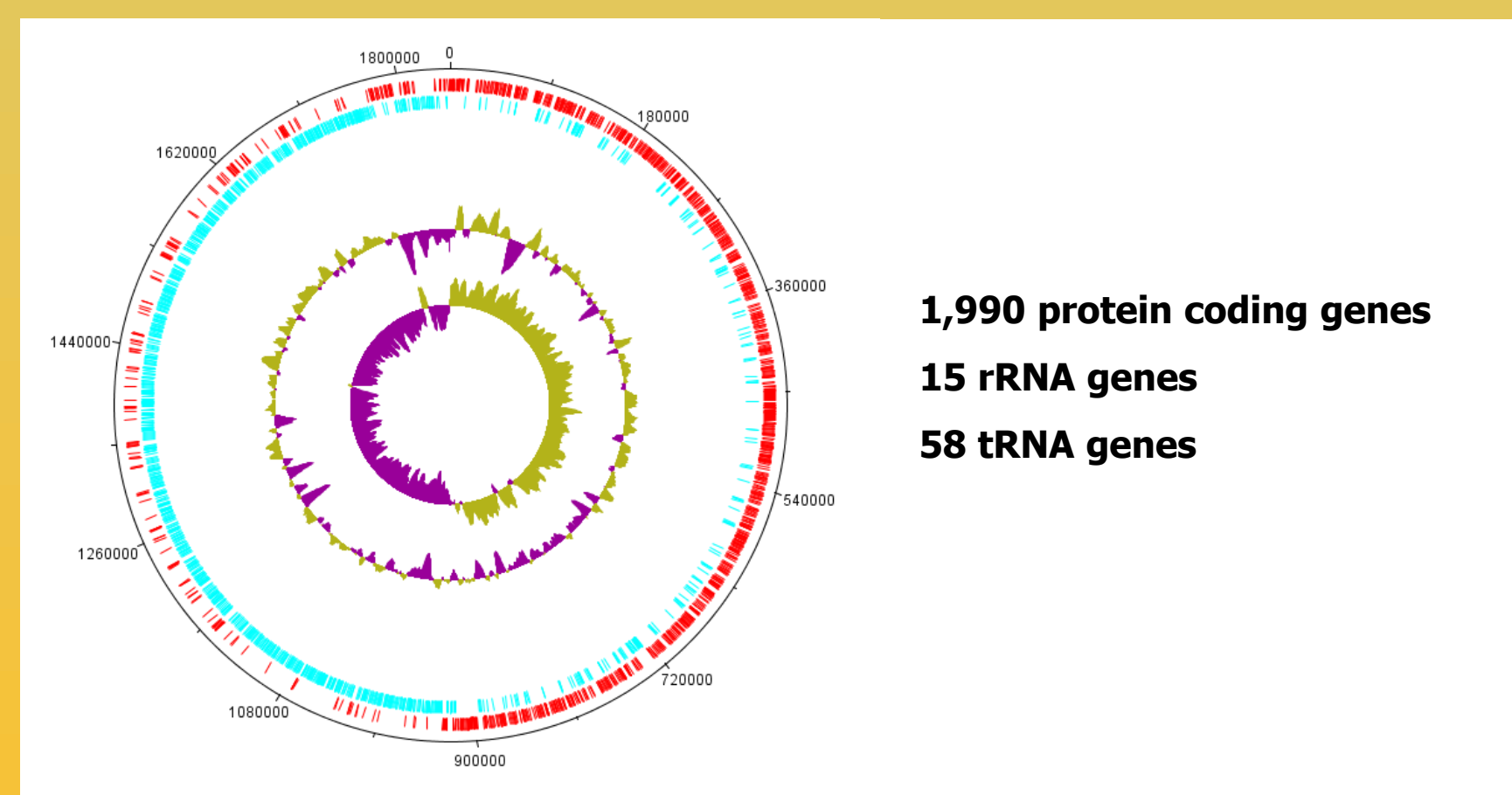


Figure 2 - A. Pairwise pan and core comparison between *Streptococcus thermophilus* ACA-DC 29 genome sequence and the nine additional complete genome sequences from dairy *Streptococcus thermophilus*. Pairs of genomes share 75% of the protein families. Homology estimation within the proteome of *Streptococcus thermophilus* ACA-DC 29 revealed that approximately 17% of the protein families had more than one member. Analysis was performed using the PanFunPro tool.

B. Pairwise proteome comparison between *Streptococcus thermophilus* ACA-DC 29 genome sequence and the nine additional complete genome sequences from dairy *Streptococcus thermophilus*. On average the proteome of *Streptococcus thermophilus* ACA-DC 29 contained 12-15% specific proteins. Analysis was performed using the PanFunPro tool.

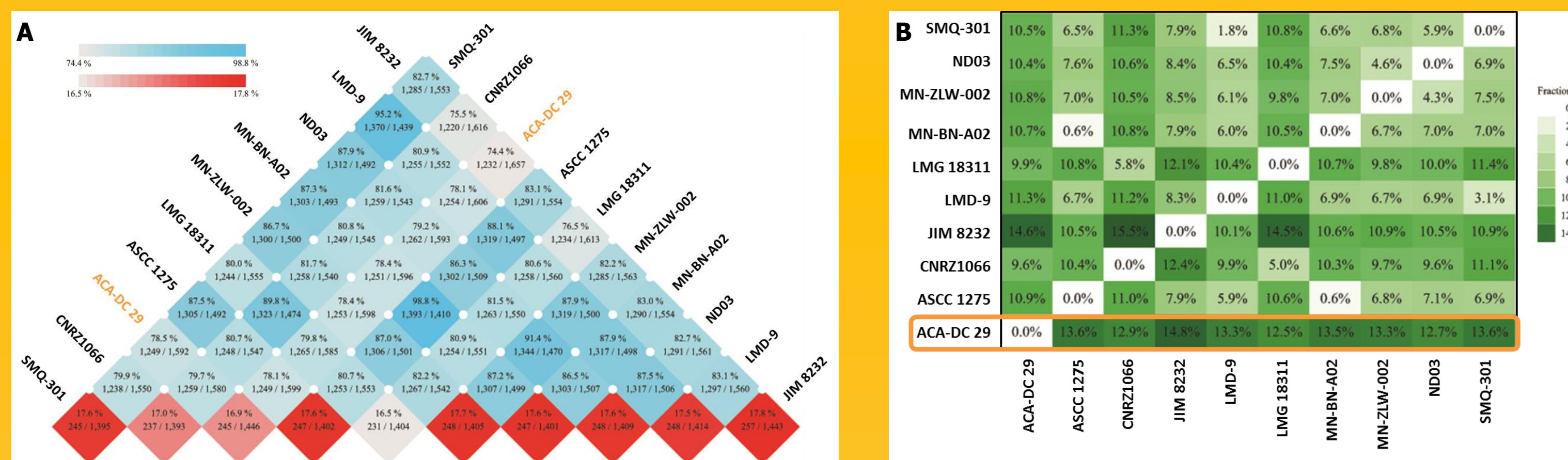


Figure 3 - Chromosome alignments of the *Streptococcus thermophilus* strains as calculated by progressiveMauve. Local collinear blocks (LCBs) of conserved sequences among the strains are represented by rectangles of the same colour. Connecting lines can be used to visualize synteny or rearrangement.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

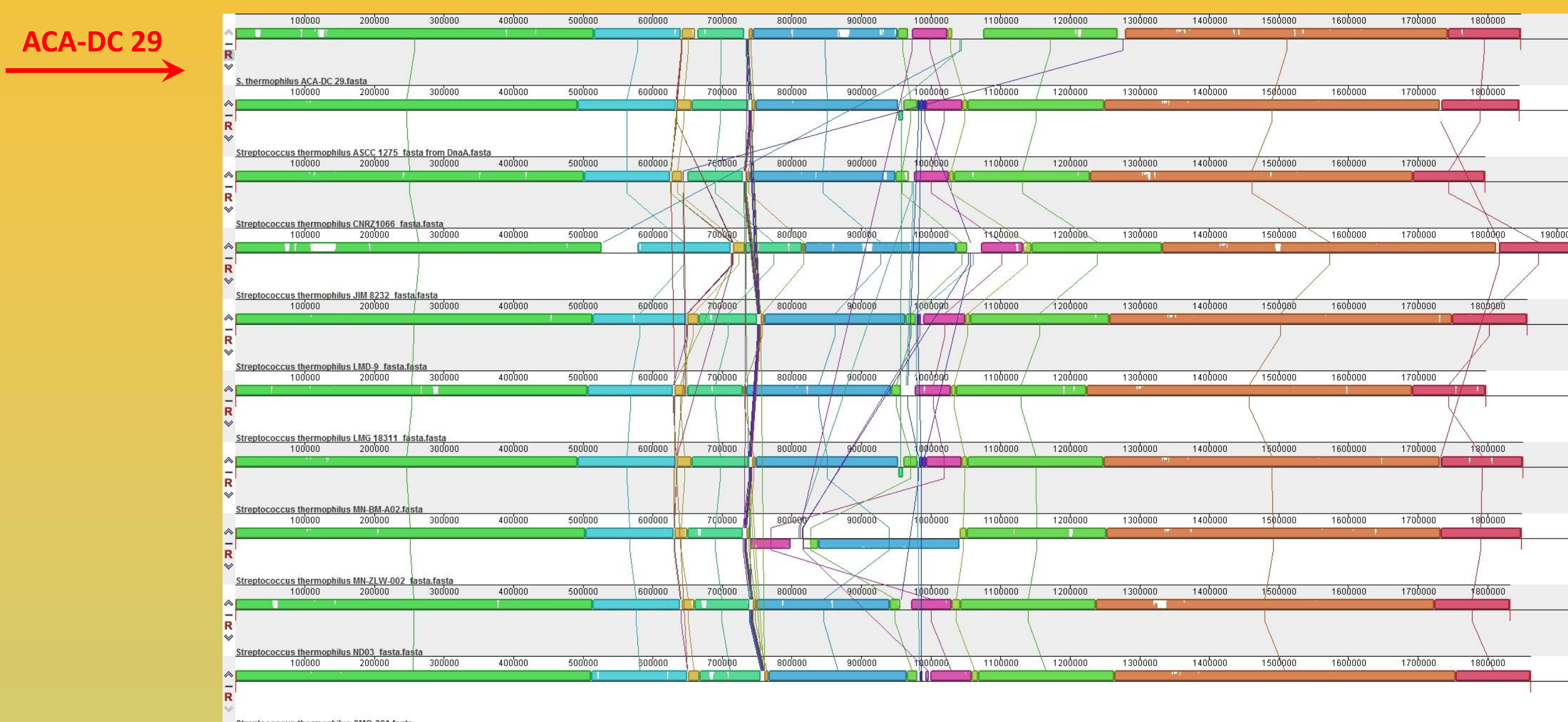


Figure 4 - Circular map of the *Streptococcus thermophilus* 29 genome. Highlighted regions correspond to genomic islands (GIs). GIs are colored within the circular maps according to the tool that predicted each one: green, orange and blue were predicted with IslandPick, SIGI-HMM and IslandPath-DIMOB, 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.

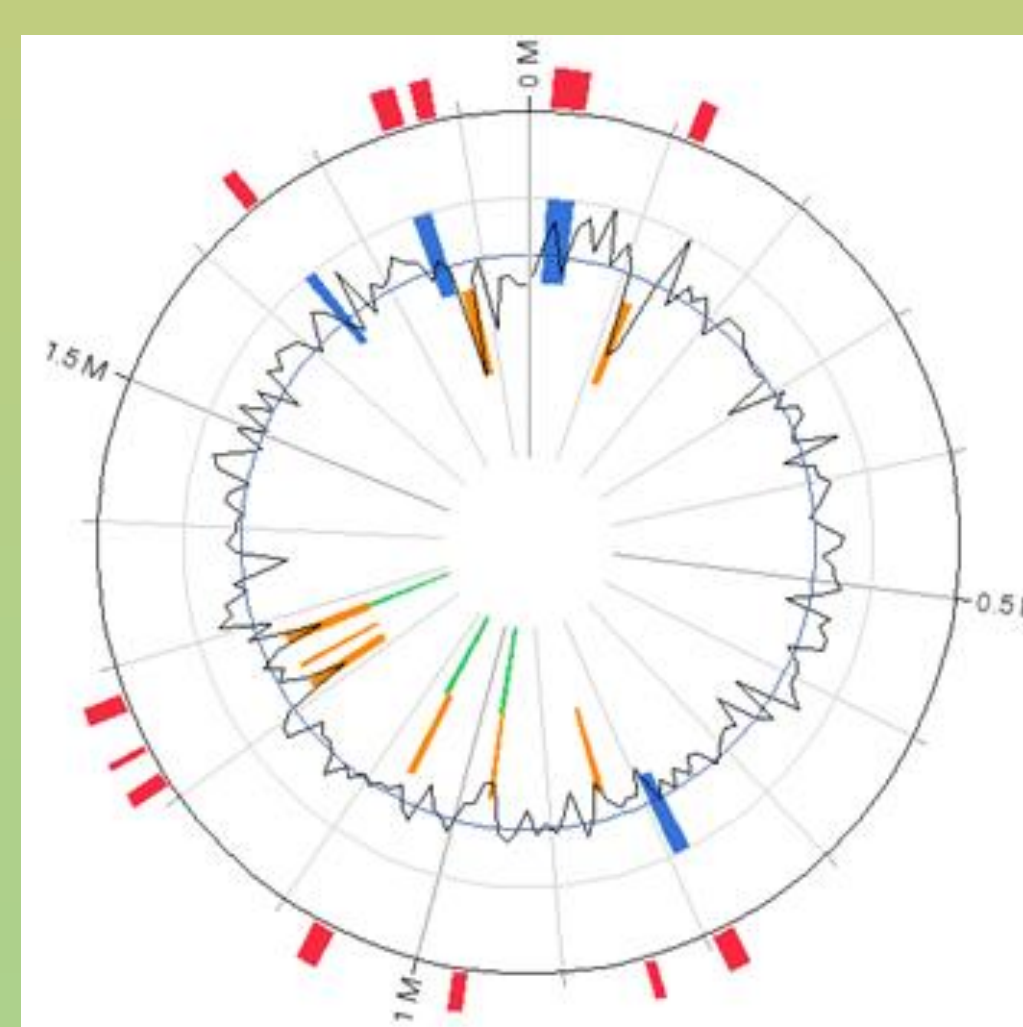


Figure 5 - A. The pangenome of the ten *Streptococcus thermophilus* strains is comprised of approximately 2,300 genes. B. The core genome, the accessory genome and the unique genes of *Streptococcus thermophilus* ACA-DC 29 strain.

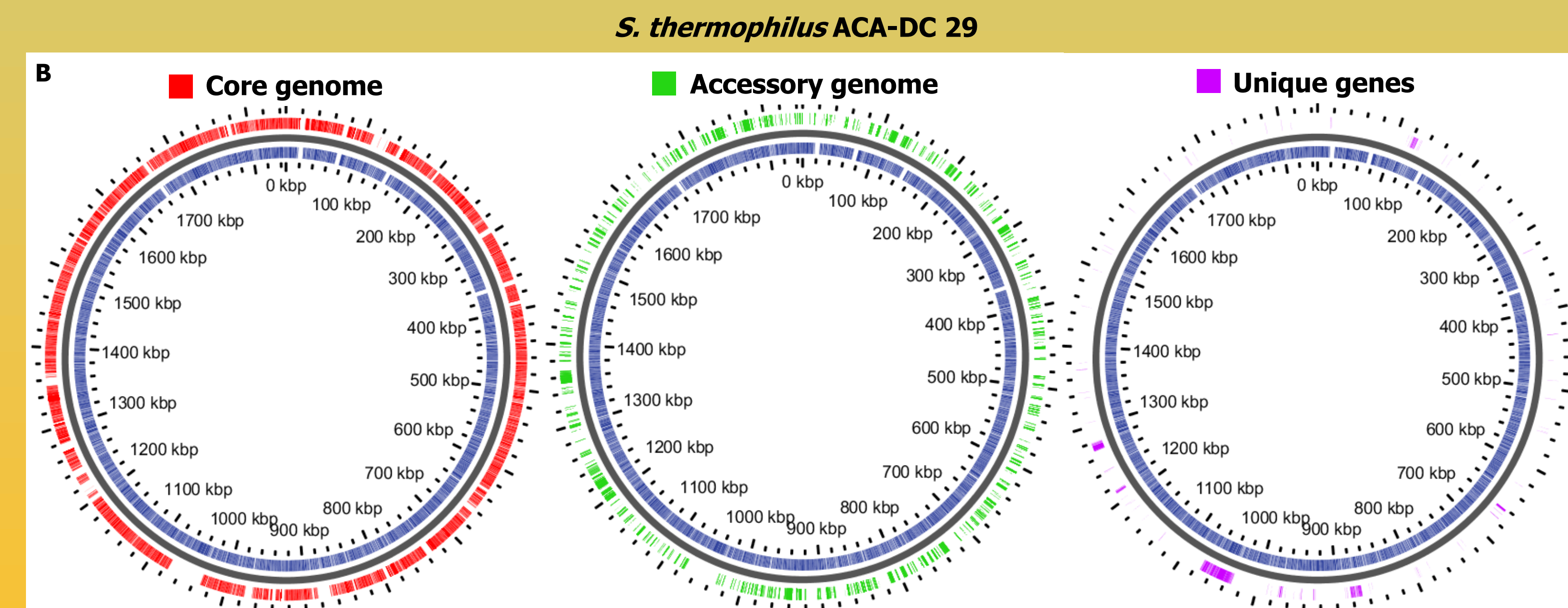
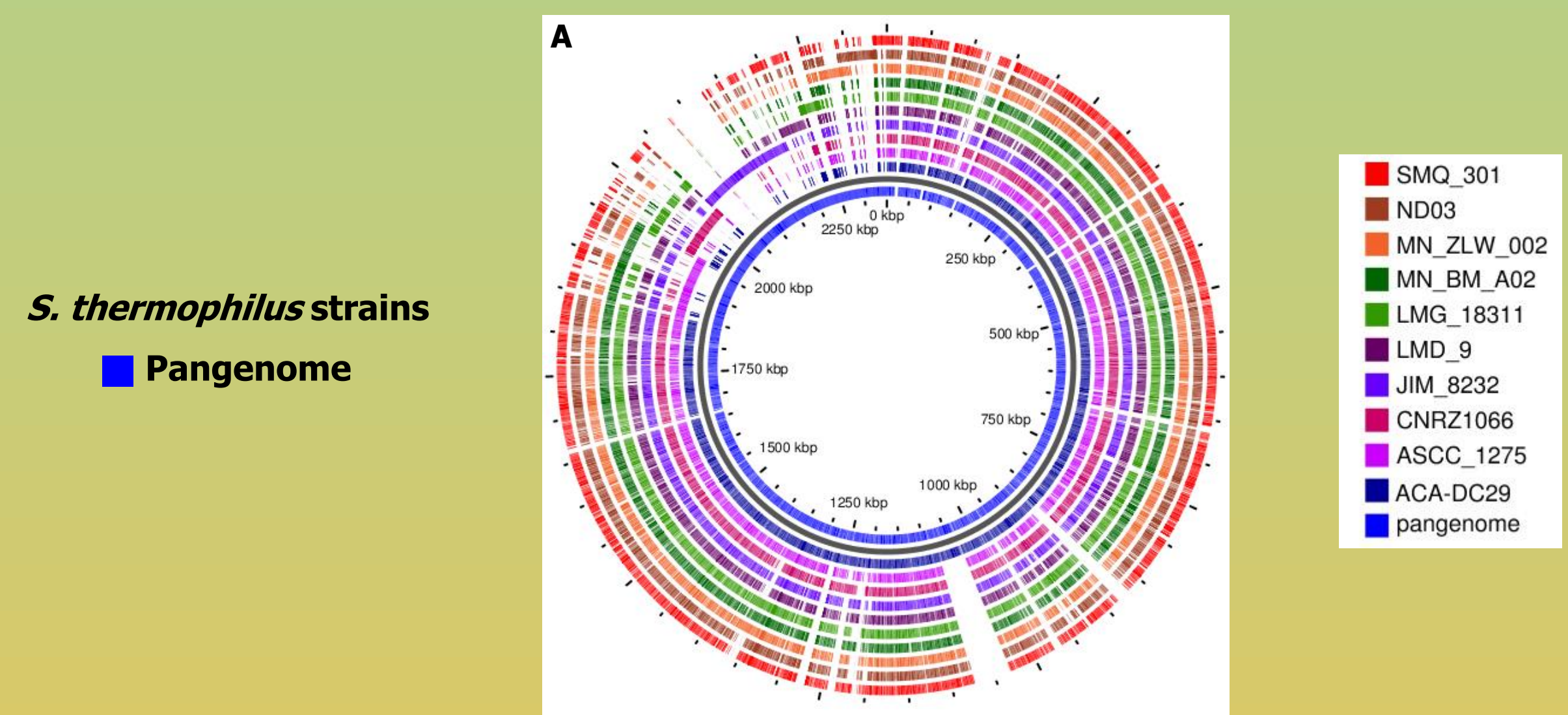


Figure 6 - A. Biological process distribution of the 250 unique genes of *S. thermophilus* ACA-DC 29. B. Top hits species distribution of the 250 unique genes of *S. thermophilus* ACA-DC 29.

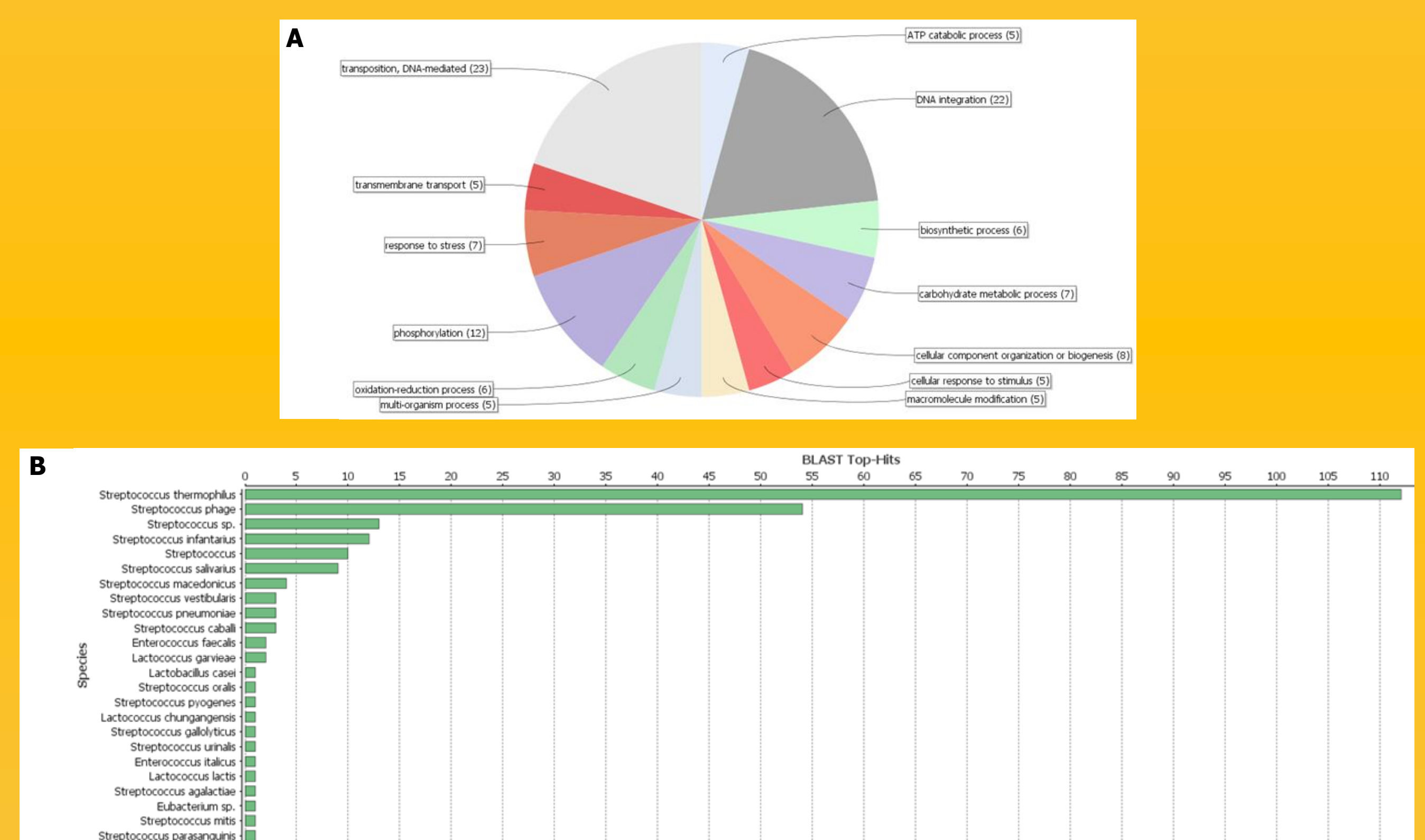


Figure 7 - Genes encoding for antimicrobial peptides in the genome sequence of *S. thermophilus* ACA-DC 29 as predicted by BAGEL3.

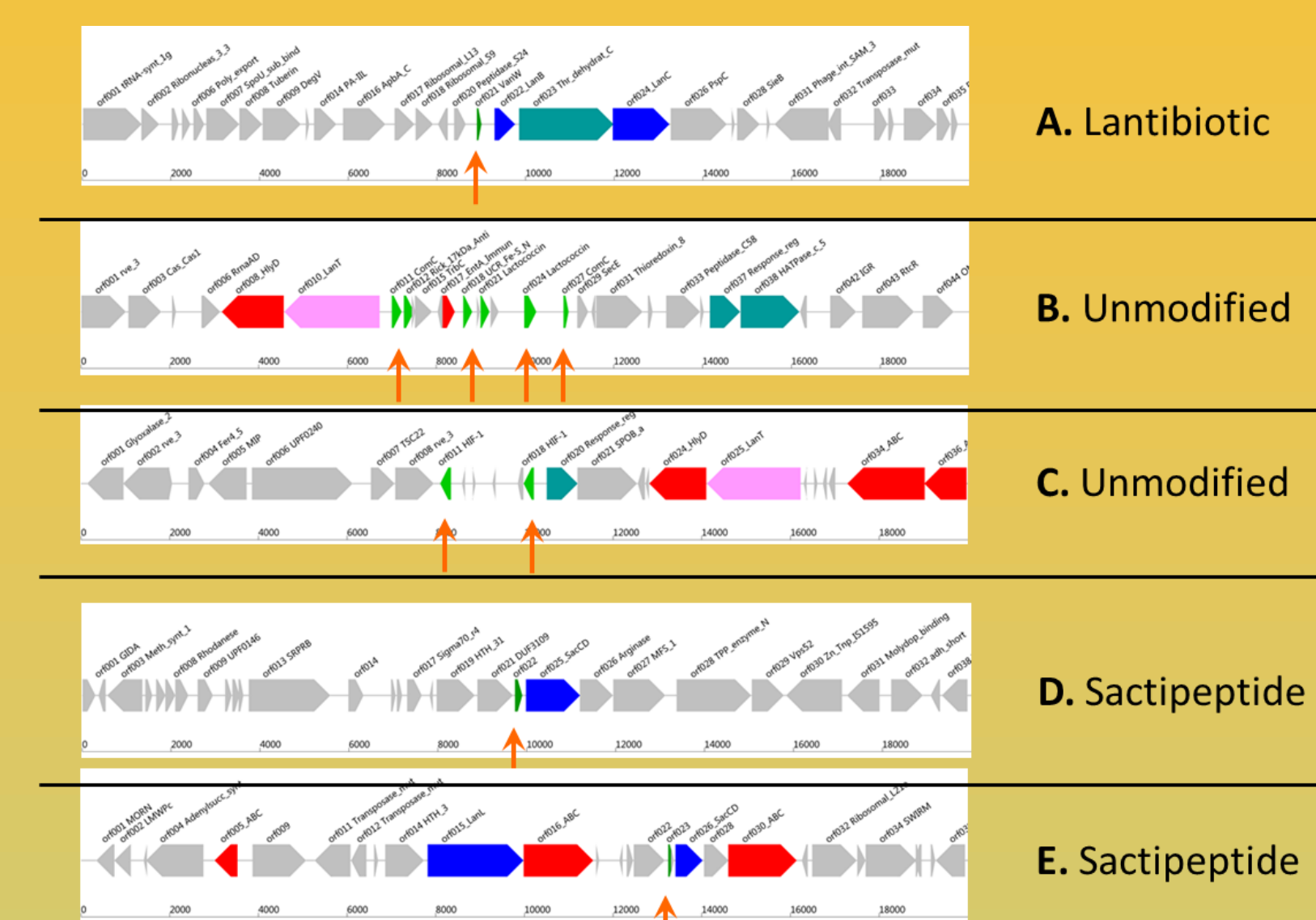
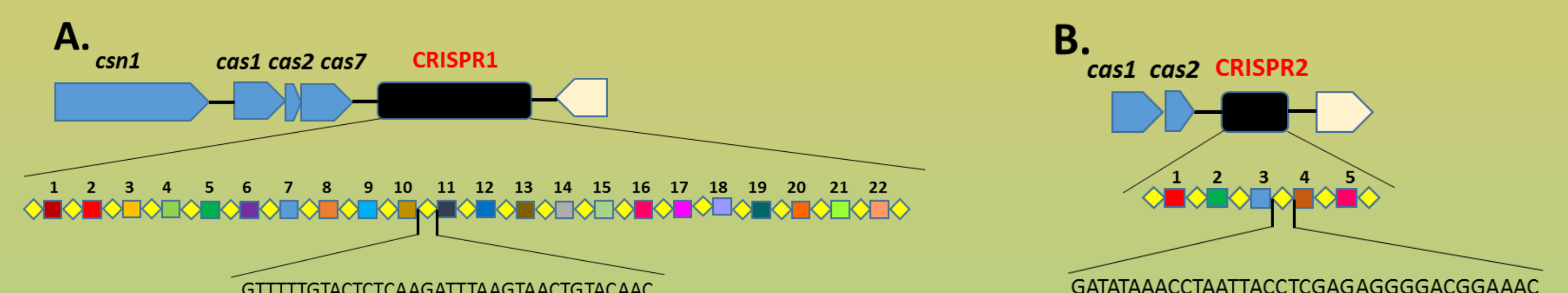


Figure 8 - CRISPR systems in the genome sequence of *S. thermophilus* ACA-DC 29 as predicted by CRISPRfinder.



Bibliography

Alexandraki V, Kazou M, Papandreou NC, Hamdrakas 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).