# Comparative genomic analysis between *Lactobacillus delbreuckii* subsp. *lactis* and *Lactobacillus delbreuckii* subsp. *bulgaricus* of dairy origin

Voula Alexandraki[1], Maria Kazou[1], Bruno Pot[2], Effie Tsakalidou[1]* and 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 (CILL), Lille, France
*e-mail: et@aua.gr

## Abstract

*Lactobacillus delbreuckii* subsp. *lactis* and *Lactobacillus delbreuckii* subsp. *bulgaricus* are lactic acid bacteria commonly exploited by the dairy industry as starter cultures, mainly for the production of cheese and yogurt. In this work, we present the comparative genomic analysis between the recently sequenced genomes of *L. lactis* ACA-DC 178 and *L. bulgaricus* ACA-DC 87, isolated from Kasseri cheese and yogurt, respectively. The genome of *L. lactis* ACA-DC 178 is larger than that of *L. bulgaricus* ACA-DC 87, containing 2,048 and 1,928 protein coding genes, respectively. Full chromosome alignments revealed strain specific differences, although a high degree of synteny between the two strains was also observed. The core genome of the two stains contained about 1,300 genes while the unique genes of the two strains were also identified. The two genomes contained several genomic islands, indicating that a number of genes have been acquired through horizontal gene transfer. We also predicted one confirmed CRISPR system in each genome, three potential antimicrobial peptides in total and also prophage sequences integrated into the genomes of the two strains. Overall, our analysis provides useful insights into the technological potential of the ACA-DC 178 and ACA-DC 87 strains.

## Results

Figure 1 - Circular maps of the genome of *L. lactis* ACA-DC 178 (A) and *L. bulgaricus* ACA-DC 87 (B). Genomic features appearing from the periphery to the center of the map: 1. Forward CDSs (cyan); 2. Reverse CDSs (red); 3. tRNA genes (blue); 4. rRNA genes (green); 5. %GC plot; 6. GC skew.
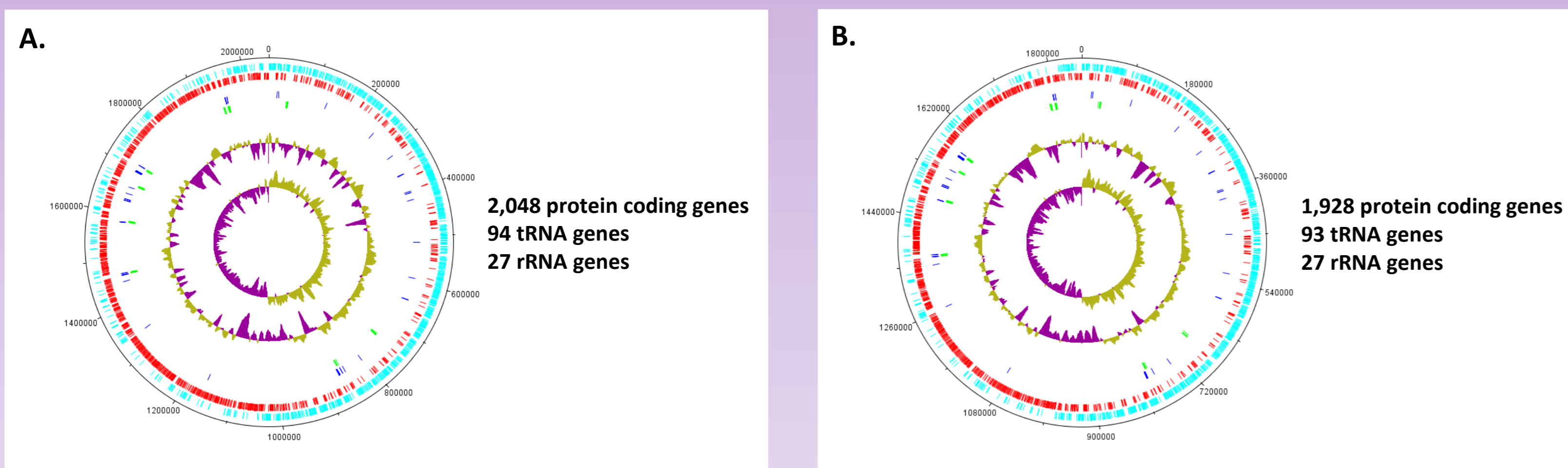


2,048 protein coding genes
94 tRNA genes
27 rRNA genes

1,928 protein coding genes
93 tRNA genes
27 rRNA genes

Figure 2 — Chromosome alignments of the strains as calculated by progressiveMauve. The *L. bulgaricus* ATCC BAA-365 and the *L. bulgaricus* ND02 strains were used as the reference genomes for the *L. bulgaricus* ACA-DC 87 and the *L. lactis* ACA-DC 178 strains, respectively. Although the ND02 strain is deposited in NCBI as *Lactobacillus delbrueckii* subsp. *bulgaricus*, Kafsi et al. 2014 suggested that the ND02 is a *Lactobacillus delbrueckii* subsp. *lactis* strain. 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.
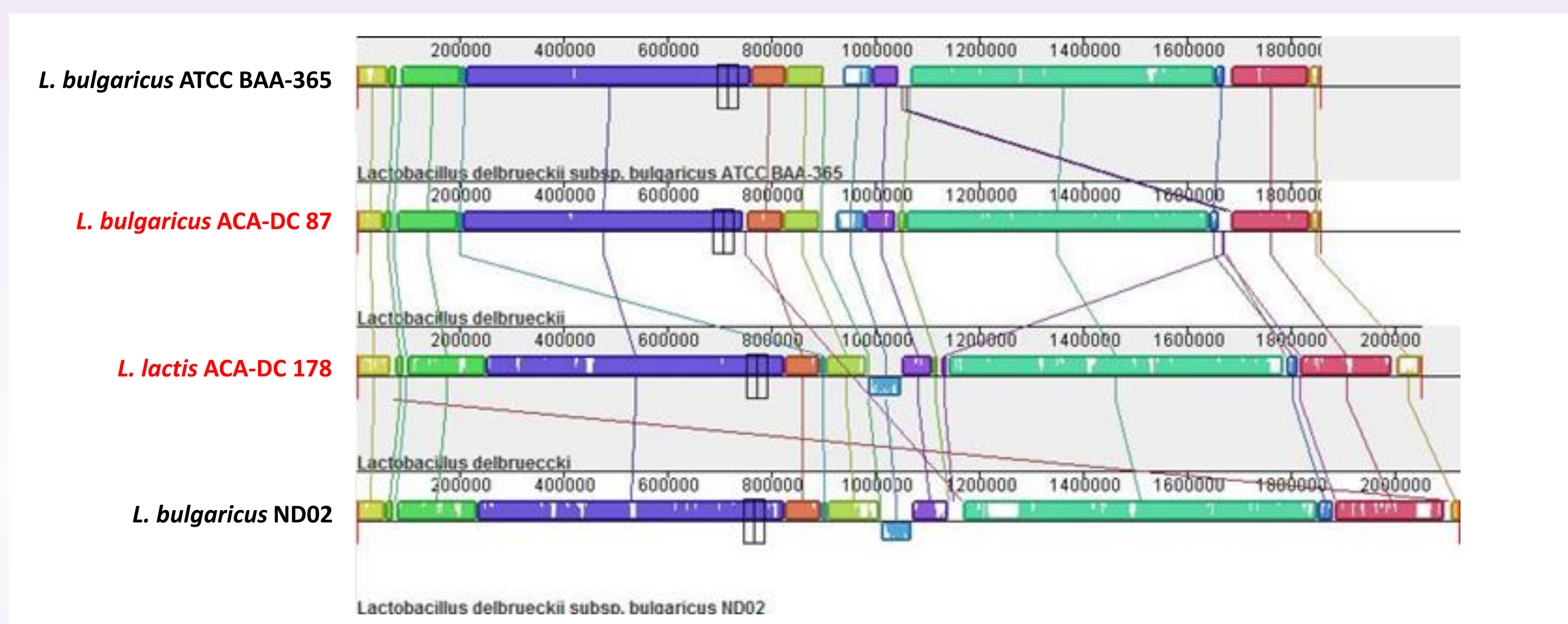


Figure 3 —The pangenome, the core genome, the accessory and the unique genes as obtained through the analysis of each set of genomes sequences with progressiveMauve. Genome sequences of *L. bulgaricus* strains (A). Genome sequences of *L. lactis* strains (B). Genome sequences of *L. bulgaricus* ACA-DC 87 and *L. lactis* ACA-DC 178 strains (C).

| A. *L. delbrueckii* subsp. *bulgaricus* | | |
|---|---|---|
| Pangenome | | 2956 |
| Core genome | | 1267 |
| Accessory genes | | 551 |
| Unique genes | ACA-DC 87 | 270 |
| | ATCC BAA-365 | 235 |
| | ATCC 11842 | 440 |
| | 2038 | 193 |

| B. *L. delbrueckii* subsp. *lactis* | | |
|---|---|---|
| Pangenome | | 2558 |
| Core genome | | 1502 |
| Unique genes | ACA-DC 178 | 546 |
| | ND02 | 510 |

| C. *L. Lactis* ACA-DC 178 vs *L. bulgaricus* ACA-DC 87 | | |
|---|---|---|
| Pangenome | | 2670 |
| Core genome | | 1306 |
| Unique genes | ACA-DC 178 | 742 |
| | ACA-DC 87 | 622 |

Figure 4 — CRISPR systems in the genome sequences of *L. lactis* ACA-DC 178 (A) and *L. bulgaricus* ACA-DC 87 (B) as predicted by CRISPRfinder



| CRISPR arrays of *L. delbrueckii* subsp. *bulgaricus* strains | | |
|---|---|---|
| Strain | CRISPR | Spacers |
| ACA-DC 87 | 1 | 11 |
| ATCC BAA-365 | 1 | 20 |
| ATCC 11842 | 1 | 40 |
| 2038 | 1 | 19 |

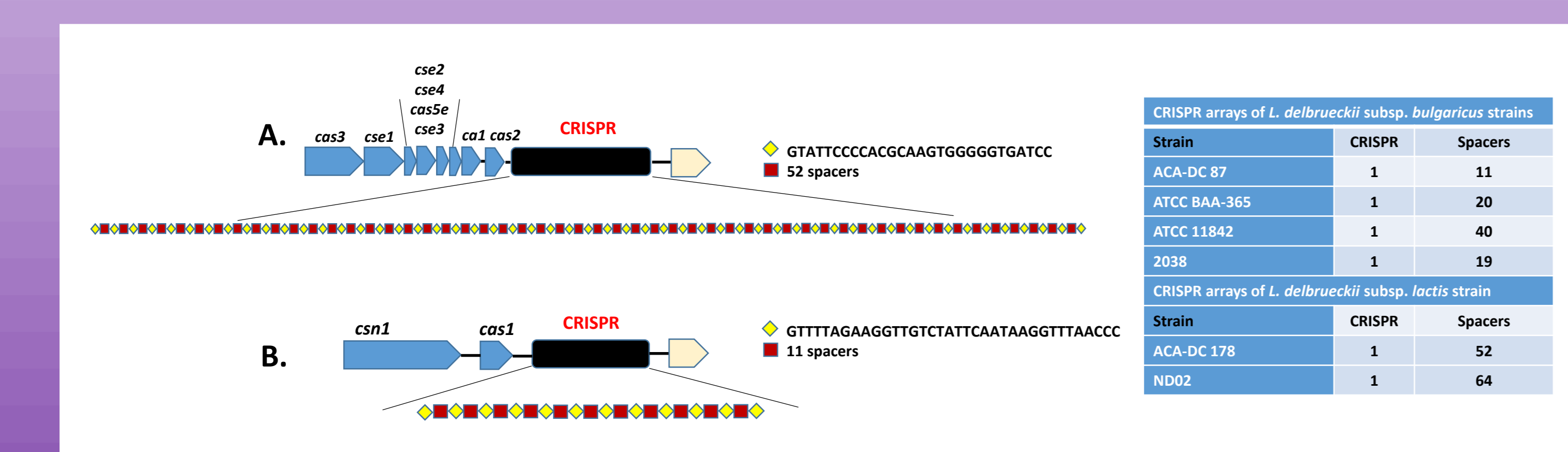| CRISPR arrays of *L. delbrueckii* subsp. *lactis* strain | | |
|---|---|---|
| Strain | CRISPR | Spacers |
| ACA-DC 178 | 1 | 52 |
| ND02 | 1 | 64 |

Figure 5 – Circular maps of *L. lactis* ACA-DC 178 (A) and *L. bulgaricus* ACA-DC 87 (B) genomes. Highlighted regions correspond to genomic islands (GIs). GIs are colored within the circular maps according to the tool that predicted each one: orange and blue were predicted with 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.
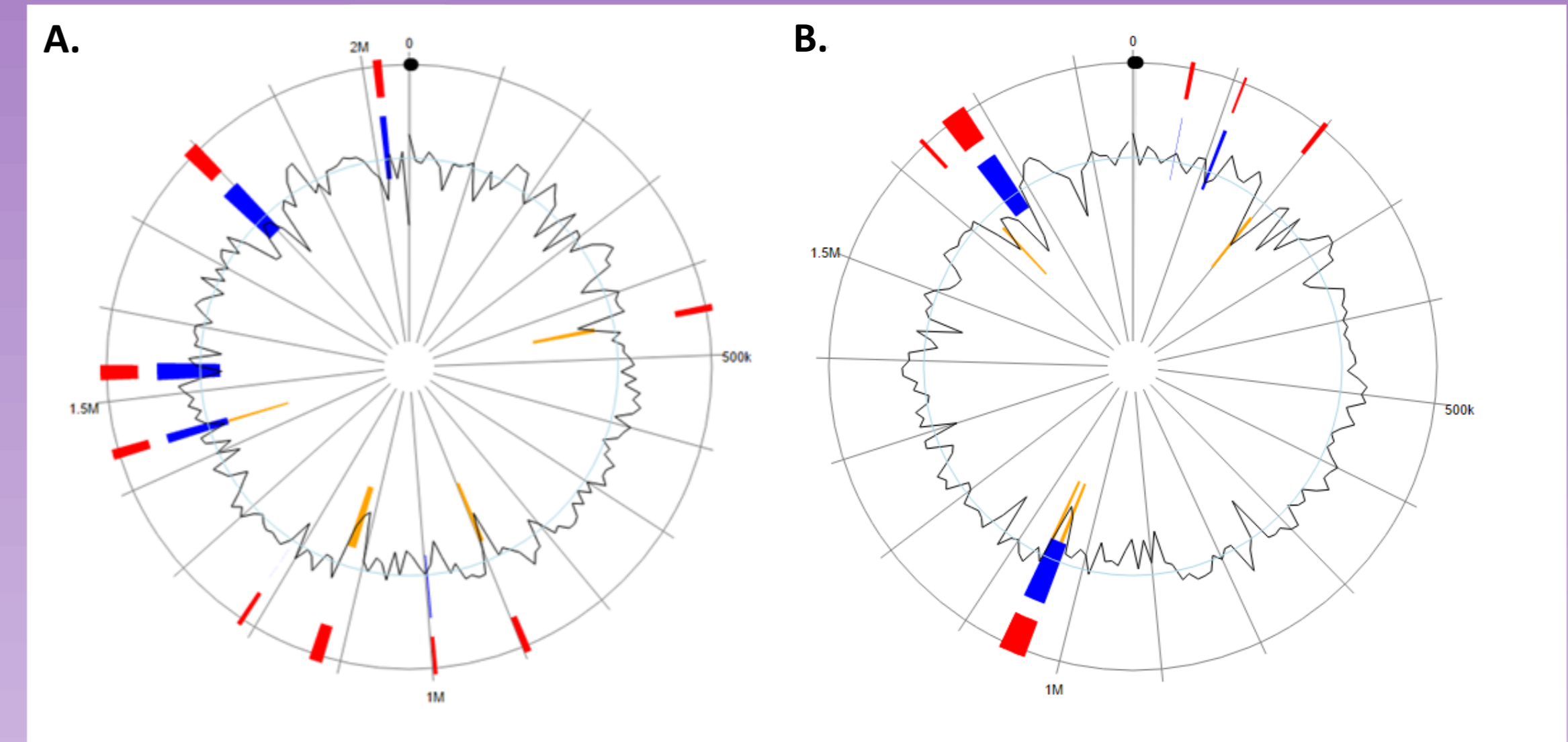


Figure 6 — Genes encoding for antimicrobial peptides in the genome sequences of *L. lactis* ACA-DC 178 (A,B) and *L. bulgaricus* ACA-DC 87 (C), as predicted by BAGEL3.
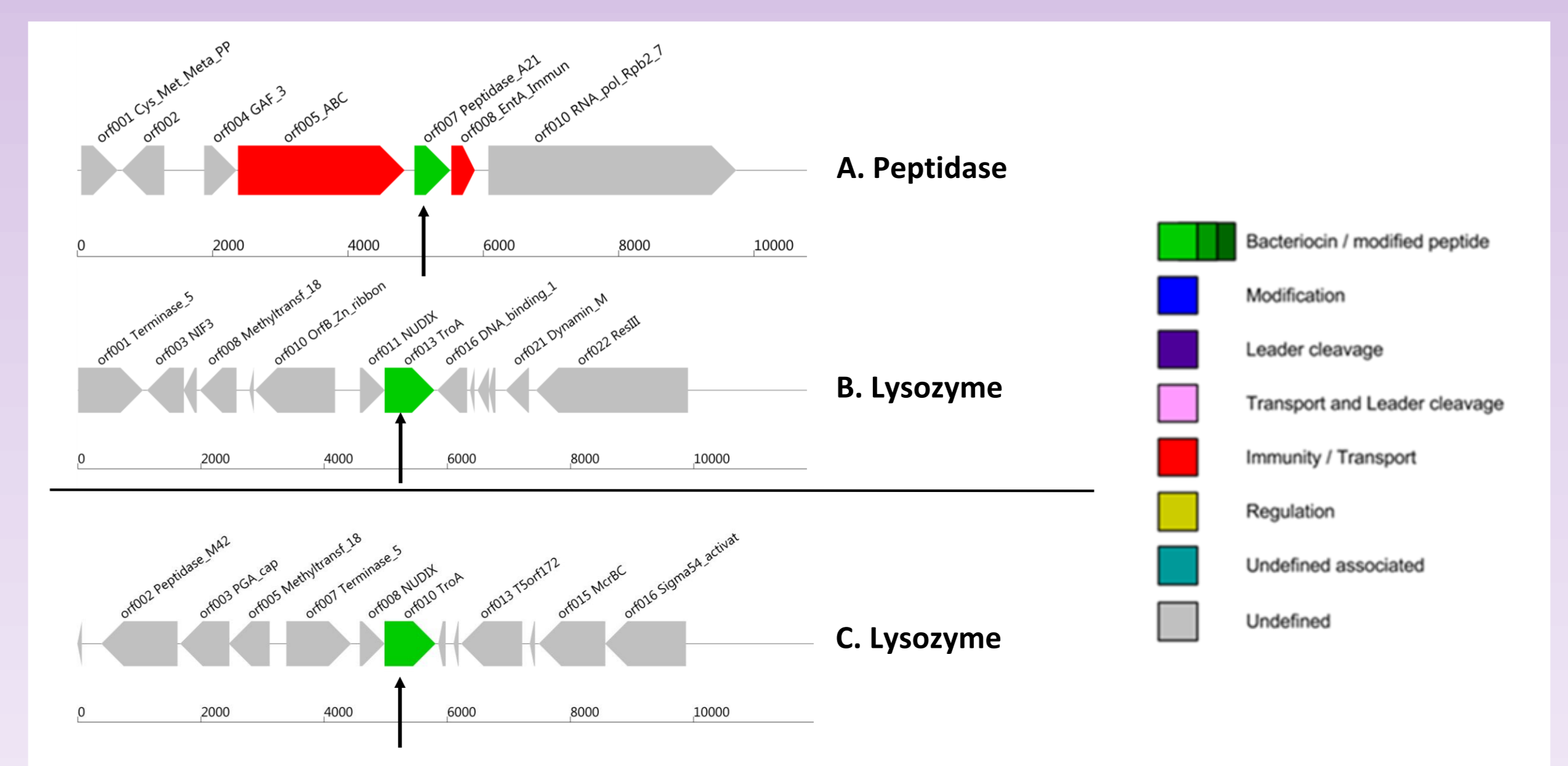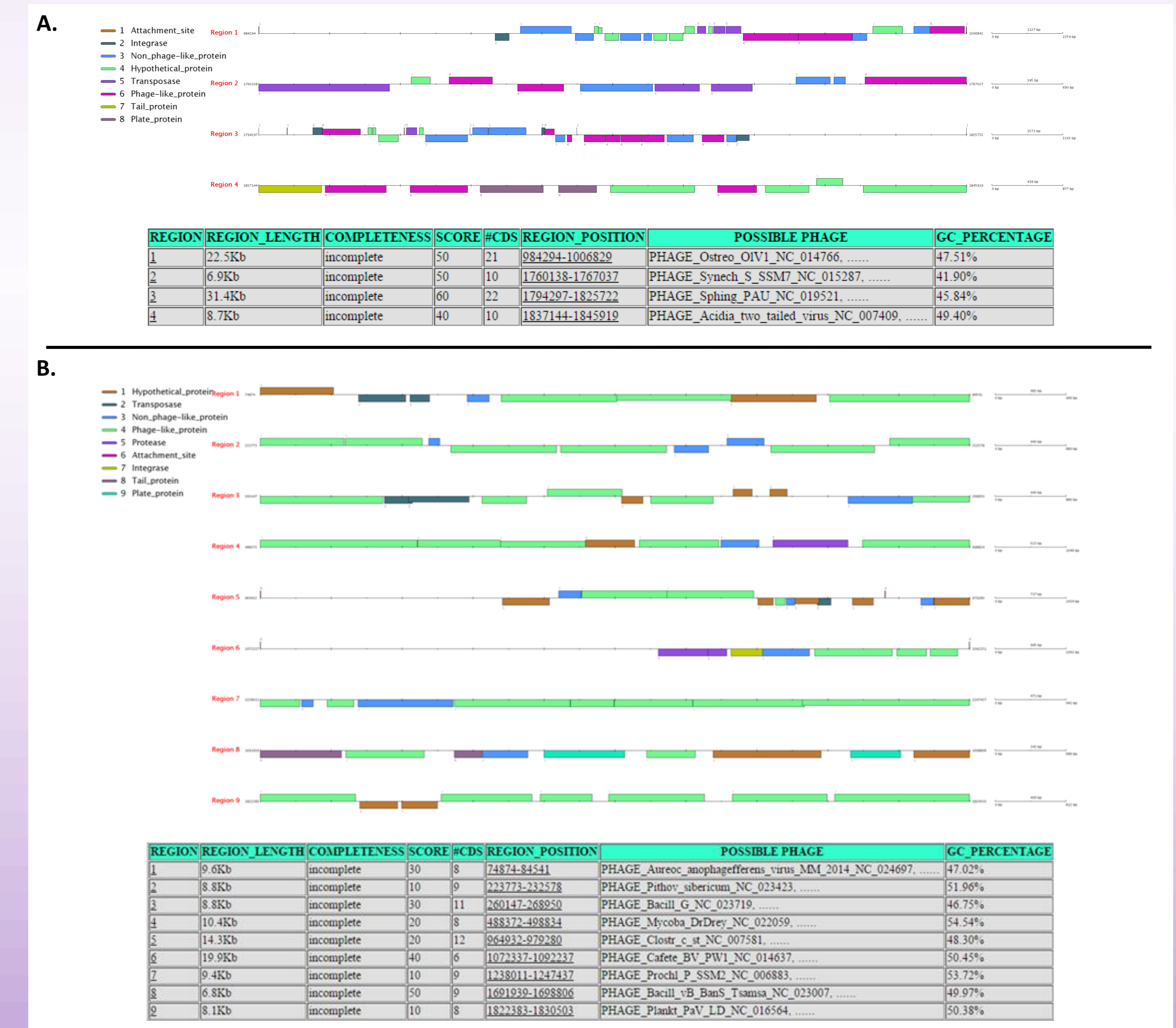


A. Peptidase
B. Lysozyme
C. Lysozyme

Bacteriocin / modified peptide
Modification
Leader cleavage
Transport and Leader cleavage
Immunity / Transport
Regulation
Undefined associated
Undefined

Figure 7 — Prophage sequences of *L. lactis* ACA-DC 178 (A) and *L. bulgaricus* ACA-DC 87 (B) as predicted by PHAST.



A.

1 Attachment_site
2 Integrase
3 Non_phage-like_protein
4 Hypothetical_protein
5 Transposase
6 Phage-like_protein
7 Tail_protein
8 Plate_protein

| REGION | REGION_LENGTH | COMPLETENESS | SCORE | #CDS | REGION_POSITION | POSSIBLE PHAGE | GC_PERCENTAGE |
|---|---|---|---|---|---|---|---|
| 1 | 22.5Kb | incomplete | 50 | 21 | 084204-1006820 | PHAGE_Ostreo_OIV1_NC_014766. .......... | 47.51% |
| 2 | 6.9Kb | incomplete | 50 | 10 | 1760138-1767037 | PHAGE_Synech_S_SSM7_NC_015287. ......... | 41.90% |
| 3 | 31.4Kb | incomplete | 60 | 22 | 1794207-1825722 | PHAGE_Sphing_PAU_NC_019521. ......... | 45.84% |
| 4 | 8.7Kb | incomplete | 40 | 10 | 1837144-1845910 | PHAGE_Acidia_two_tailed_virus_NC_007409. ....... | 49.40% |

B.

1 Hypothetical_protein
2 Transposase
3 Non_phage-like_protein
4 Phage-like_protein
5 Protease
6 Attachment_site
7 Integrase
8 Tail_protein
9 Plate_protein

| REGION | REGION_LENGTH | COMPLETENESS | SCORE | #CDS | REGION_POSITION | POSSIBLE PHAGE | GC_PERCENTAGE |
|---|---|---|---|---|---|---|---|
| 1 | 9.6Kb | incomplete | 30 | 8 | 74874-84541 | PHAGE_Aureoc_anophagefferens_virus_MM_2014_NC_024697. ... | 47.02% |
| 2 | 8.8Kb | incomplete | 10 | 9 | 225773-232578 | PHAGE_Pithov_ubericum_NC_023423. ......... | 51.96% |
| 3 | 8.8Kb | incomplete | 30 | 11 | 260147-268950 | PHAGE_Bacill_G_NC_023719. ......... | 46.75% |
| 4 | 10.4Kb | incomplete | 20 | 8 | 488372-498834 | PHAGE_Mycoba_DrDerp_NC_022059. ......... | 54.54% |
| 5 | 14.3Kb | incomplete | 20 | 12 | 564932-579280 | PHAGE_Clostr_c_st_NC_007581. | 48.50% |
| 6 | 19.9Kb | incomplete | 40 | 6 | 1072337-1092237 | PHAGE_Cafete_BV_PW1_NC_014637. ......... | 50.45% |
| 7 | 9.4Kb | incomplete | 10 | 9 | 1238011-1247437 | PHAGE_Prochl_P_SSM2_NC_006883. ......... | 53.72% |
| 8 | 6.8Kb | incomplete | 50 | 9 | 1691030-1698806 | PHAGE_Bacill_vB_BanS_Tsamsa_NC_023007. ......... | 49.97% |
| 9 | 8.1Kb | incomplete | 10 | 8 | 1822383-1830503 | PHAGE_Plankt_PaV_LD_NC_016564. ......... | 50.38% |

## Bibliography

Hela El Kafsi, Johan Binesse, Valentin Loux, Julien Buratti, Samira Boudebbouze, Rozenn Dervyn, Sean Kennedy, Nathalie Galleron, Benoit Quinquis, Jean-Michel Batto, Bouziane Moumen, Emmanuelle Maguin and Maarten van de Guchte (2014). *Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus*: a chronicle of evolution in action. BMC Genomics, 15:407.

## Acknowledgments