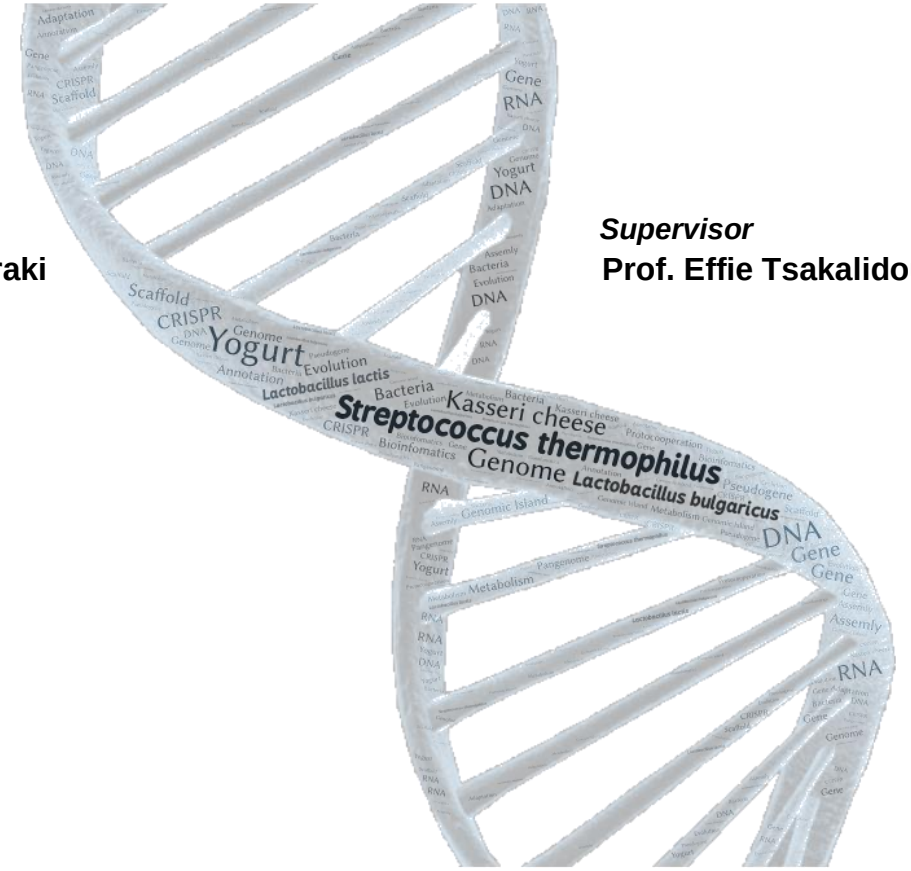
ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
AGRICULTURAL UNIVERSITY OF ATHENS

Department of Food Science and Human Nutrition

Laboratory of Dairy Research

Doctoral thesis

Whole genome sequencing and characterization of the lactic acid bacteria *Streptococcus thermophilus, Lactobacillus delbrueckii* subsp*. bulgaricus* and *Lactobacillus delbrueckii* subsp*. lactis*. Physiological, evolutionary and technological implications



Author
Stavroula Alexandraki

Supervisor
Prof. Effie Tsakalidou

Athens 2020

AGRICULTURAL UNIVERSITY OF ATHENS
DEPARTMENT OF FOOD SCIENCE AND HUMAN NUTRITION
Laboratory of Dairy Research

DOCTORAL THESIS

"Whole genome sequencing and characterization of the lactic acid bacteria *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Lactobacillus delbrueckii* subsp. *lactis*. Physiological, evolutionary and technological implications"

Αλληλούχηση και χαρακτηρισμός των γονιδιωμάτων των οξυγαλακτικών βακτηρίων *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* και *Lactobacillus delbrueckii* subsp. *lactis*. Φυσιολογικές, εξελικτικές και τεχνολογικές προεκτάσεις.

Stavroula Alexandraki

Members of the Advisory Committee:
Tsakalidou Effie, Professor Agricultural University of Athens (supervisor)
Nychas Georgios-Ioannis, Professor Agricultural University of Athens
Papadimitriou Konstantinos, Assistant Professor University of Peloponnese

Members of the Examination Committee
Vorgias Constantinos, Professor National and Kapodistrian University of Athens
Skandamis Panagiotis, Associate Professor Agricultural University of Athens
Panagou Efstathios, Associate Professor Agricultural University of Athens
Moschopoulou Ekaterini, Assistant Professor Agricultural University of Athens

Whole genome sequencing and characterization of the lactic acid bacteria *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Lactobacillus delbrueckii* subsp. *lactis*. Physiological, evolutionary and technological implications

Department of Food Science and Human Nutrition
Laboratory of Dairy Research

Abstract

Lactic acid bacteria (LAB) have been traditionally related to the production of fermented food and feed and are generally considered beneficial microorganisms, with several strains exhibiting probiotic potency. However, some LAB genera like *Streptococcus* and *Enterococcus*, include pathogenic species for both human and animals. Thus, a thorough comprehension of taxonomy, metabolism and molecular biology of LAB is essential to fully take advantage of their technological, nutritional and health-promoting traits, while avoiding at the same time potential risks. Nowadays, this is feasible due to the advancement of sequencing technologies, which have enhanced the generation of high quality genome sequences. Furthermore, the continuous development of bioinformatics tools enables the mining of biological data through *in silico* analysis. The informative value of such genomic scrutiny is notable, since it yields insights about strain-dependent technological and probiotic features, as well as bacterial diversity and niche-related adaptability.

In the present thesis, the complete genome sequences of three LAB strains of the ACA-DC collection of the Laboratory of Dairy Research of the Agricultural University of Athens, namely *Streptococcus thermophilus* ACA-DC 2, *Lactobacillus delbrueckii* subsp. *bulgaricus* ACA-DC 87 and *Lactobacillus delbrueckii* subsp. *lactis* 178, were produced by next-generation sequencing (NGS) technologies and studied *in silico* by employing various bioinformatics tools, aiming at deciphering their biological and technological potential. Strains ACA-DC 2 and ACA-DC 87 were isolated from traditional naturally fermented Greek yogurt while strain ACA-DC 178 was isolated from naturally fermented Greek Kasseri cheese.

With chromosome size of 1.73 Mbp, *S. thermophilus* ACA-DC 2 is the strain with the smallest genome among the strains of the species with complete sequenced genomes. Its genome carries 1,850 genes, with 1,556 encoding proteins, while 224 were identified as putative pseudogenes. Taking into consideration the lack of pathogenicity along with the presence of

numerous pseudogenes, the strain has probably evolved through genome decay towards its adaption to the milk ecosystem. Clusters of Orthologous Groups (COG) functional classification of proteins showed that 1,327 of them (approximately 85%) were assigned to at least one COG category, with the most abundant being related to amino acid transport and metabolism (E: 10%), and that approximately 28.5% of the proteins do not have any described function. Among the studied technological-related genomic features of ACA-DC 2 were one complete lactose-galactose operon and one exopolysaccharide (EPS) gene cluster, implicated in lactose metabolism and EPS production, respectively. In addition, several proteolytic enzymes were predicted, although the cell-envelope proteinase (CEP) PrtS was absent. Stress response genes and four putative antimicrobial peptides were also identified. The presence of one putative Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-CRISPR associated (Cas) system was confirmed. However, the activity of the Cas proteins may be compromised, since the CRISPR array owns only one spacer. Finally, the existence of four putative restriction-modification (R-M) systems may account for the inadequacy of the CRISPR-Cas system.

In order to assess the evolutionary, biological and technological traits of *S. thermophilus*, comparative genomic analysis among 23 strains of the species with complete genomes was performed. Comprehensive pangenomic analysis showed that the species pan genome include 2,516 genes, while the core genome comprises an important number of conserved genes, namely 1,082 genes. Moreover, 997 accessory genes along with numerous unique genes in several strains were also identified. Analysis also determined that the species pan genome will be shortly closed. Based on whole genome phylogenetic and average nucleotide identity (ANI) analyses, *S. thermophilus* strains, except for strain NCTC12958^{T}, were grouped in two separate clusters, one including strains with genomes above 1.83 Mbp (i.e. cluster A), and the rest strains with smaller genomes (i.e. cluster B). This differentiation may be attributed to distinct gene gain or loss events that took place during the evolution of the species. Moreover, certain *S. thermophilus* strains formed subgroups within the aforementioned clusters. More specifically, cluster A included four subgroups, namely I (ASCC 1275, DGCC 7710, KLDS SM, MN-BM-A02, and ND07), II (MN-BM-A01 and MN-ZLW-002), III (LMD-9 and SMQ-301), and IV (APC151 and ND03), while cluster B contained only one subgroup (CNRZ1066, CS8, EPS, and S9). The occurrence of clusters and subgroups among *S. thermophilus* strains suggests the presence of lineages within the species.

Thorough investigation concerning the allocation and/or the architecture of numerous genomic features, such as protease PrtS, histidine biosynthetic cluster, EPS loci, CRISPR-Cas systems, R-M systems and genomic islands (GIs), corroborated this observation to a certain extent. Moreover, several genetic loci implicated in essential metabolic processes, like lactose and urea metabolism, amino acid biosynthesis and amino acid and peptide transport, were found to be conserved within the species.

Analysis of the *Lb. delbrueckii* subsp. *bulgaricus* ACA-DC 87 chromosomal sequence (1.86 Mbp) revealed the presence of 1,993 genes, 1,644 protein-coding genes and 229 putative pseudogenes. Approximately 87% of the protein-coding genes (1,284 proteins) were assigned to at least one COG functional category, with the majority of proteins being allocated primarily to the category of translation, ribosomal structure and biogenesis (J: 8.5%), and secondarily to the categories of amino acid transport and metabolism (E: 7.8%) and replication, recombination, and repair (L: 7.4%). One CRISPR array with a size of 761 bp, carrying 11 spacers and two distinct loci related to EPS biosynthesis were predicted. Finally, 12 GIs holding 196 genes in total, several of which encode CRISPR-associated proteins, subunits of R-M systems, and proteins implicated in EPS biosynthesis, were also identified.

Regarding the genome sequence of *Lb. delbrueckii* subsp. *lactis* ACA-DC 178 (2.05 Mbp), a total of 2,112 genes were identified, including 1,752 protein-coding genes and 239 putative pseudogenes. COG annotation showed that about 80% of the proteins (i.e. 1,417) were assigned to at least one functional category, with the majority being distributed among the categories of replication, recombination and repair (L: 10.6%), translation, ribosomal structure and biogenesis (J: 7.5%), and amino acid transport and metabolism (E: 7.2%). One large CRISPR array of 3,197 bp containing 52 spacers, several of which are identical to phage sequences having hosts in strains of the genus *Lactobacillus*, was identified. Furthermore, two EPS biosynthetic gene clusters and 14 integrated GIs with a total of 159 genes, encoding proteins associated with EPS biosynthesis, amino acid transport and subunits of R-M systems, were also found. These findings clearly demonstrate that whole genome analysis and comparative genomics facilitated the identification of several genetic features associated with important technological traits, highlighting the significance of the application of genomics in food-related microorganisms.

Scientific area: Natural Sciences, Biological Sciences

Key words: Lactic acid bacteria; Milk; Yogurt; Cheese; Lactobacillus; Streptococcus; Genomics; Bioinformatics; Technological properties; Pan genome, CRISPR

Αλληλούχηση και χαρακτηρισμός των γονιδιωμάτων των οξυγαλακτικών βακτηρίων *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* και *Lactobacillus delbrueckii* subsp. *lactis*. Φυσιολογικές, εξελικτικές και τεχνολογικές προεκτάσεις.

Τμήμα Επιστήμης Τροφίμων και Διατροφής του Ανθρώπου
Εργαστήριο Γαλακτοκομίας

Περίληψη

Τα οξυγαλακτικά βακτήρια εφαρμόζονται στην παραγωγή τροφίμων ζύμωσης και ζωοτροφών και γενικά θεωρούνται ωφέλιμοι μικροοργανισμοί, με αρκετά στελέχη να παρουσιάζουν προβιοτικές ιδιότητες. Ωστόσο, ορισμένα γένη των οξυγαλακτικών βακτηρίων, όπως τα γένη *Streptococcus* και *Enterococcus*, περιλαμβάνουν παθογόνα είδη τόσο για τον άνθρωπο όσο και για τα ζώα. Επομένως η κατανόηση της ταξινόμησης, του μεταβολισμού και της μοριακής βιολογίας των οξυγαλακτικών βακτηρίων είναι μείζονος σημασίας για την μέγιστη αξιοποίηση των τεχνολογικών, διατροφικών και προβιοτικών ιδιοτήτων τους, αποφεύγοντας ταυτόχρονα πιθανούς κινδύνους. Σήμερα, αυτό είναι εφικτό λόγω της βελτίωσης των τεχνικών αλληλούχησης, οι οποίες επιτρέπουν την παραγωγή γονιδιωμάτων υψηλής ποιότητας. Επιπλέον, η συνεχής ανάπτυξη εργαλείων βιοπληροφορικής επιτρέπει την ανεύρεση βιολογικών δεδομένων μέσω της *in silico* ανάλυσης των αλληλουχιών. Η πληροφοριακή δύναμη μιας τέτοιας γονιδιωματικής ανάλυσης είναι αξιοσημείωτη, καθώς παρέχει γνώσεις για τα τεχνολογικά και προβιοτικά χαρακτηριστικά, καθώς και για την βακτηριακή ποικιλομορφία και την οικολογική προσαρμογή των βακτηριακών στελεχών.

Στην παρούσα διδακτορική διατριβή, μελετήσαμε το τεχνολογικό δυναμικό τριών οξυγαλακτικών βακτηρίων της συλλογής μικροοργανισμών ACA-DC του Εργαστηρίου Γαλακτοκομίας του Γεωπονικού Πανεπιστημίου Αθηνών, χρησιμοποιώντας σύγχρονες τεχνικές αλληλούχησης και προηγμένα εργαλεία βιοπληροφορικής. Πιο συγκεκριμένα μελετήθηκαν τα στελέχη *Streptococcus thermophilus* ACA-DC 2 και *Lactobacillus delbrueckii* subsp. *bulgaricus* ACA-DC 87, οι οποίοι απομονώθηκαν από παραδοσιακό γιαούρτι, καθώς και το στέλεχος *Lactobacillus delbrueckii* subsp. *lactis* 178, το οποίο απομονώθηκε από παραδοσιακό Κασέρι.

Με χρωμόσωμα περίπου 1.73 Mbp, ο *S. thermophilus* ACA-DC 2 είναι το στέλεχος με το μικρότερο γονιδίωμα μεταξύ των στελεχών του είδους, για τα οποία υπάρχουν ολοκληρωμένες αλληλουχίες. Σε αυτό βρέθηκαν 1,850 γονίδια, εκ των οποίων τα 1,556 κωδικοποιούν πρωτεΐνες,

καθώς και 224 πιθανά ψευδογονίδια. Η απουσία χαρακτηριστικών παθογένειας καθώς και ο μεγάλος αριθμός ψευδογονιδίων υποδεικνύουν εκφυλισμό του γονιδιώματος και υποστηρίζουν την προσαρμογή του στελέχους στο περιβάλλον του γάλακτος. Μελέτη της λειτουργίας των πρωτεϊνών με τη βάση δεδομένων Clusters of Orthologous Groups (COG) έδειξε ότι 1,327 πρωτεΐνες (περίπου 85%) ανήκουν σε μία τουλάχιστον λειτουργική κατηγορία COG, με επικρατέστερη την κατηγορία για την μεταφορά και τον μεταβολισμό αμινοξέων (E: 10%), ενώ περίπου 28.5% των πρωτεϊνών δεν έχουν κάποια χαρακτηρισμένη λειτουργία.

Στο στέλεχος ACA-DC 2 ταυτοποιήθηκαν γενωμικά χαρακτηριστικά που σχετίζονται με σημαντικές τεχνολογικές ιδιότητες και περιλαμβάνουν τα οπερόνια για τον καταβολισμό της λακτόζης και της γαλακτόζης, το οπερόνιο για τη βιοσύνθεση εξωπολυσακχαριτών, τα πρωτεολυτικά ένζυμα, τους μηχανισμούς απόκρισης σε συνθήκες στρες καθώς και γονίδια για την παραγωγή αντιμικροβιακών ενώσεων. Το στέλεχος δεν κωδικοποιεί την πρωτεάση PrtS. Επίσης, επιβεβαιώθηκε η ύπαρξη ενός πιθανού Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-CRISPR associated (Cas) συστήματος. Ωστόσο η ενεργότητα των Cas πρωτεϊνών πιθανόν να είναι χαμηλή, αφού το στέλεχος διαθέτει μόνο έναν spacer στη συστοιχία CRISPR. Τέλος, η ύπαρξη τεσσάρων πιθανών restriction-modification (R-M) συστημάτων, ενδεχομένως να εξισορροπεί την ανεπάρκεια του συστήματος CRISPR-Cas.

Προκειμένου να αξιολογήσουμε τα εξελικτικά, βιολογικά και τεχνολογικά χαρακτηριστικά του *S. thermophilus*, πραγματοποιήθηκε συγκριτική γονιδιωματική ανάλυση μεταξύ 23 στελεχών του είδους με ολοκληρωμένα γονιδιώματα. Το πανγονιδίωμα του είδους περιλαμβάνει 2,516 γονίδια, εκ των οποίων τα 1,082 ανήκουν στο συντηρημένο γονιδίωμα ενώ 997 ανήκουν στο συμπληρωματικό γονιδίωμα. Επίσης προσδιορίστηκαν μοναδικά γονίδια σε αρκετά από τα στελέχη που μελετήθηκαν. Η ανάλυση έδειξε ότι το πανγονιδίωμα του είδους είναι ακόμα ανοιχτό, σύντομα όμως θα μπορούσε να χαρακτηριστεί ως κλειστό.

Η φυλογενετική ανάλυση των στελεχών του *S. thermophilus* βάσει του συντηρημένου γονιδιώματος καθώς και του ποσοστού ταυτόσημων νουκλεοτιδίων (average nucleotide identity, ANI), έδειξε την ύπαρξη δύο διακριτών Clusters, το Cluster A που περιλαμβάνει 15 στελέχη με μέγεθος γονιδιωμάτων μεγαλύτερο από 1.83 Mbp, και το Cluster B που περιλαμβάνει 7 στελέχη με μικρότερο μέγεθος γονιδιωμάτων. Το στέλεχος NCTC12958^T δεν ανήκει σε κανένα από τα δύο Clusters. Αυτή η διαφοροποίηση πιθανόν να οφείλεται σε διακριτά

γεγονότα απόκτησης ή απώλειας γονιδίων κατά την εξελικτική πορεία του είδους. Σε αυτά τα δύο Clusters εντοπίστηκαν υποομάδες στελεχών. Πιο συγκεκριμένα το Cluster A περιλαμβάνει τέσσερις υποομάδες στελεχών, τις I (ASCC 1275, DGCC 7710, KLDS SM, MN-BM-A02, and ND07), II (MN-BM-A01 and MN-ZLW-002), III (LMD-9 and SMQ-301), and IV (APC151 and ND03), ενώ το Cluster B περιλαμβάνει μόνο μία υποομάδα στελεχών (CNRZ1066, CS8, EPS, and S9). Η ύπαρξη Clusters και υποομάδων για τα στελέχη του *S. thermophilus* υποδεικνύει την ύπαρξη γενεαλογικών γραμμών του είδους. Διεξοδική μελέτη της κατανομής και/ή της αρχιτεκτονικής ποικίλλων γενωμικών χαρακτηριστικών, όπως η πρωτεάση PrtS, τα γονίδια για τη βιοσύνθεση της ιστιδίνης, οι γενετικές περιοχές για τη βιοσύνθεση εξωπολυσακχαριτών, τα συστήματα CRISPR-Cas και R-M και οι γονιδιωματικές νήσοι, επιβεβαιώνουν ως ένα βαθμό την παραπάνω παρατήρηση. Επιπλέον, πολλά γονίδια που σχετίζονται με βασικές μεταβολικές λειτουργίες, όπως τον μεταβολισμό της λακτόζης και της ουρίας, τη βιοσύνθεση των αμινοξέων και το πρωτεολυτικό σύστημα, βρέθηκαν συντηρημένα σε μεγάλο βαθμό σε όλα τα στελέχη του είδους.

Ο *Lb. delbrueckii* subsp. *bulgaricus* ACA-DC 87 έχει χρωμόσωμα με μέγεθος περίπου 1.86 Mbp. Σε αυτό βρέθηκαν 1,993 γονίδια εκ των οποίων τα 1,644 κωδικοποιούν πρωτεΐνες. Επίσης ταυτοποιήθηκαν 229 πιθανά ψευδογονίδια. Περίπου 87% των πρωτεϊνών (1,284 πρωτεΐνες) ανήκουν σε μία τουλάχιστον λειτουργική κατηγορία COG, με την πλειοψηφία των πρωτεϊνών να εντοπίζονται κυρίως στις κατηγορίες της μετάφρασης, της ριβοσωμικής δομής και της βιογένεσης (J: 8.5%), της μεταφοράς και του μεταβολισμού αμινοξέων (E: 7.8%) και της αναπαραγωγής, του ανασυνδυασμού και της επιδιόρθωσης (L: 7.4%). Επίσης, ταυτοποιήθηκε μία συστοιχία CRISPR με μέγεθος 761 bp και 11 spacers καθώς και δύο διακριτές γενετικές περιοχές για τη βιοσύνθεση εξωπολυσακχαριτών. Τέλος, βρέθηκαν 12 γονιδιωματικές νήσοι, οι οποίες περιλαμβάνουν συνολικά 196 γονίδια, πολλά από τα οποία κωδικοποιούν Cas πρωτεΐνες, πρωτεΐνες των συστημάτων R-M καθώς και πρωτεΐνες για την παραγωγή εξωπολυσακχαριτών.

Ο *Lb. delbrueckii* subsp. lactis ACA-DC 178 έχει χρωμόσωμα με μέγεθος περίπου 2.05 Mbp. Σε αυτό βρέθηκαν 2,112 γονίδια, εκ των οποίων τα 1,752 κωδικοποιούν πρωτεΐνες. Επίσης ταυτοποιήθηκαν 239 πιθανά ψευδογονίδια. Λειτουργική ανάλυση των πρωτεϊνών με τη βάση δεδομένων COG αποκάλυψε ότι περίπου το 80% των πρωτεϊνών (1,417 πρωτεΐνες) ανήκουν σε μία τουλάχιστον λειτουργική κατηγορία COG, με την πλειοψηφία αυτών να εντοπίζονται στις

κατηγορίες της αναπαραγωγής, του ανασυνδυασμού και της επιδιόρθωσης (L: 10.6%), της μετάφρασης, της ριβοσωμικής δομής και της βιογένεσης (J: 7.5%) και της μεταφοράς και του μεταβολισμού αμινοξέων (E: 7.2%). Επίσης ταυτοποιήθηκε μία μεγάλη συστοιχία CRISPR με μέγεθος 3,197 bp και 52 spacers, πολλοί από τους οποίους αντιστοιχούν σε φάγους του γένους *Lactobacillus*. Επιπλέον, βρέθηκαν δύο διακριτές περιοχές για τη βιοσύνθεση εξωπολυσακχαριτών καθώς και 14 γονιδιωματικές νήσοι, οι οποίες περιλαμβάνουν συνολικά 159 γονίδια. Κάποια από αυτά τα γονίδια σχετίζονται με τη μεταφορά αμινοξέων, τα συστήματα R-M καθώς και την παραγωγή εξωπολυσακχαριτών. Αυτά τα ευρήματα καταδεικνύουν σαφώς ότι η ανάλυση ολόκληρων γονιδιωμάτων καθώς και η συγκριτική γονιδιωματική ανάλυση διευκόλυναν την ταυτοποίηση πολλών γενετικών χαρακτηριστικών που σχετίζονται με σημαντικές τεχνολογικές ιδιότητες, τονίζοντας τη σημασία της εφαρμογής της γονιδιωματικής ανάλυσης σε μικροοργανισμούς που σχετίζονται με τα τρόφιμα.

Επιστημονική περιοχή: Φυσικές Επιστήμες, Βιολογικές Επιστήμες

Λέξεις κλειδιά: Οξυγαλακτικά βακτήρια; Γάλα; Γιαούρτι; Τυρί; Lactobacillus; Streptococcus; Γονιδιωματική; Βιοπληροφορική; Τεχνολογικές Ιδιότητες; Πανγονιδίωμα; CRISPR

To my beloved father

who anticipated the completion of this thesis

I miss you every day

Whole genome sequencing and characterization of the lactic acid bacteria *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Lactobacillus delbrueckii* subsp. *lactis*. Physiological, evolutionary and technological implications.

Thesis committee

Thesis supervisor

Professor Effie Tsakalidou
Department of Food Science and Human Nutrition
Agricultural University of Athens, Greece

Thesis co-supervisors

Professor Georgios-Ioannis Nychas
Department of Food Science and Human Nutrition
Agricultural University of Athens, Greece

Assistant Professor Konstantinos Papadimitriou
Department of Food Science and Technology
University of Peloponnese, Kalamata, Greece

Assistant Professor Konstantinos Papadimitriou was assigned as thesis co-supervisor in the place of Professor Constantinos E. Vorgias, Department of Biochemistry and Molecular Biology National and Kapodistrian University of Athens [decision of the Department of Food Science and Human Nutrition Assembly 557th/06.02.2020, according to the provisions of Article 39, Law No 4485/2017]. Professor Constantinos E. Vorgias was assigned as thesis co-supervisor in the place of Professor Emeritus Stavros-Ioannis Hamodrakas, Department of Cell Biology and Biophysics, National and Kapodistrian University of Athens, who meanwhile retired [decision of the Department of Food Science and Human Nutrition Assembly 537th/05.07.2018, according to the provisions of Article 39, Law No 4485/2017].

Other members

Professor Constantinos E. Vorgias
Department of Biochemistry and Molecular Biology
National and Kapodistrian University of Athens, Greece

Associate Professor Panagiotis Skandamis
Department of Food Science and Human Nutrition
Agricultural University of Athens, Greece

Associate Professor Efstathios Z. Panagou
Department of Food Science and Human Nutrition
Agricultural University of Athens, Greece

Assistant Professor Ekaterini E. Moschopoulou
Department of Food Science and Human Nutrition
Agricultural University of Athens, Greece

Acknowledgments

It is with immense gratitude that I acknowledge the contribution of my supervisor, Professor Effie Tsakalidou, not only during my doctoral studies over the last years but also to the completion of this dissertation. This thesis would have remained unfinished had it not been her advices and patience when difficulties arose. Her faith in me when I needed it the most gave me the boost to move forward. I am beholden to her for the firm guidance, the constructive criticism and the dedicated support throughout the years I worked with her. The experience and the knowledge I gained during my studies under her supervision is inestimable.

I am sincerely grateful and indebted to my advisor Assistant Professor Konstantinos Papadimitriou, whose expertise was invaluable. I would like to thank him for his guidance and assistance during all these years. Moreover, I am thankful for everything he taught me, for his advises and encouragement to improve my scientific skills and for the incessantly and tirelessly education he provided for me during my doctoral studies. Foremost, I would like to thank him for his patience and tireless efforts to support me with my thesis during the last years.

I would like to wholeheartedly thank Professor George-John Nychas, Professor Emeritus Stavros-Ioannis Hamodrakas and Professor Constantinos E. Vorgias for being my advisers during my doctoral studies. Furthermore, I would like to thank all the members of the exams committee, Associate Professor Panagiotis Skandamis, Associate Professor Efstathios Panagou and Assistant Professor Ekaterini Moschopoulou for devoting their time in reviewing my thesis.

I would also like to thank Dr. Maria Kazou, with whom we started as PhD partners and eventually we became good friends, supporting each other over the years. I am delighted for everything we got through together during this academic endeavor, especially our conversations and cheery moments.

Moreover, I would like to sincerely thank Dr. Rania Anastasiou, Dr. Georgia Zoumpopoulou, Dr. Marina Georgalaki and Dr. Eugenia Manolopoulou for our fruitful collaboration. Their friendly manners and their willingness to assist me throughout these years made me feel part of their team and helped me enjoy my experience during my stay in the laboratory. I would also like to genuinely thank all the people in the Laboratory of Dairy Research for helping me whenever it was needed.

Last, but not least I would like to express my deep and sincere gratitude to my parents, my father Panagiotis and my mother Athanasia, for their endless love and understanding and the constant support towards the completion of my thesis, along with my brother Dimitris, who always has my back. Most of all I am grateful to my father, who left us so unexpectedly. He was always there for me and he supported me selflessly in every way I needed. I cannot put into words what he meant to me. I want only to thank him from the bottom of my heart for the way he raised me. I could not wish for a nicer father and his absence is greatly felt. I will always cherish you.

List of publications

A. Publications within this thesis

I. Journal papers

1. **Alexandraki V.**, Kazou M., Blom J., Pot B., Papadimitriou K., Tsakalidou E. (2019) Comparative genomics of *Streptococcus thermophilus* support important traits concerning the evolution, biology and technological properties of the species. *Front Microbiol, 10: 2916.* doi: 10.3389/fmicb.2019.02916
2. **Alexandraki V.**, Kazou M., Pot B., Tsakalidou E., Papadimitriou K. (2019) Whole-genome sequence data and analysis of *Lactobacillus delbrueckii* subsp. *lactis* ACA-DC 178 isolated from Greek Kasseri cheese. *Data Brief*, *25:104282.* doi: 10.1016/j.dib.2019.104282
3. **Alexandraki V.**, Kazou M., Pot B., Tsakalidou E., Papadimitriou K. (2017) Complete Genome Sequence of the Yogurt Isolate *Lactobacillus delbrueckii* subsp. *bulgaricus* ACA-DC 87. *Genome Announc, 5(34) pii: e00868-17.* doi: 10.1128/genomeA.00868-17
4. **Alexandraki V.**, Kazou M., Blom J., Pot B., Tsakalidou E., Papadimitriou K. (2017) The complete genome sequence of the yogurt isolate *Streptococcus thermophilus* ACA-DC 2. *Stand Genomic Sci, 12:18.* doi: 10.1186/s40793-017-0227-5

II. International conference papers

1. **Alexandraki V.**, Kazou M., Pot B., Tsakalidou E., Papadimitriou K. (2016) *In silico* assessment of the technological potential of the dairy *Streptococcus thermophilus* ACA-DC 2 through genome analysis and comparative genomics. *12th International Association for Food Protection (IAFP), 11-13 May 2016, Athens, Greece* (poster)
2. **Alexandraki V.**, Kazou M., Pot B., Tsakalidou E., Papadimitriou K. (2015) Comparative genomic analysis between *Lactobacillus delbreuckii* subsp. *lactis* and *Lactobacillus delbreuckii* subsp. *bulgaricus* of dairy origin. *World Food Day - Mediterranean Food and Nutrition, 16-17 October, Portorož, Slovenia* (poster)
3. Kazou M., **Alexandraki V.**, Renault P., Pot B., Tsakalidou E., Papadimitriou K. (2015) Adaptation of *Streptococcus macedonicus* and *Streptococcus thermophilus* in milk. Common strategies, distinct ways. *World Food Day - Mediterranean Food and Nutrition, 16-17 October, Portorož, Slovenia* (poster)
4. **Alexandraki V.**, Kazou M., Renault P., Pot B., Tsakalidou E., Papadimitriou K. (2015) Characterization of the dairy *Streptococcus thermophilus* ACA-DC 29 strain through comparative genomics. *FEMS 6th Congress of European Microbiologists, 7-11 June, Maastricht, the Netherlands* (poster)

III. National conference papers

1. **Alexandraki V.**, Kazou M., Pot B., Tsakalidou E., Papadimitriou K. (2015) Comparative genomic analysis of *Lactobacillus delbreuckii* subsp. *lactis* and *Lactobacillus delbreuckii* subsp. *bulgaricus* dairy strains. *66th Conference of Hellenic Society for Biochemistry and Molecular Biology (HSBMB), 11-13 December, Athens, Greece* (poster)
2. **Alexandraki V.**, Kazou M., Papandreou N.C., Hamodrakas S.J., Pot B., Tsakalidou E., Papadimitriou K. (2014) Comparative genomics among dairy strains of *Streptococcus thermophilus*. *65th Conference of Hellenic Society for Biochemistry and Molecular Biology, 28-30 November, Thessaloniki, Greece* (poster)
3. **Alexandraki V.**, Sarafianou A., Kazou M., Anastasiou R., Papandreou N.C., Hamodrakas S. J., Pot B., Renault P., Tsakalidou E., Papadimitriou K. (2014) Comparative genomics among dairy strains of *Streptococcus thermophilus* and *Streptococcus macedonicus*. *9th Conference of the Hellenic Society of Computational Biology and Bionformatics HSCBB2014, 10-12 October, Athens, Greece* (oral presentation)

B. Other publications during this thesis

I. Journal papers

1. Papadimitriou K., Anastasiou R., Georgalaki M., Bounenni R., Paximadaki A., Charmpi C., **Alexandraki V.**, Kazou M., Tsakalidou E. (2020) The microbial ecosystem of industrial and homemade Feta cheese revealed through 16S rDNA and ITS amplicon metagenomics as well as shotgun metagenomics (In preparation)
2. Zoumpopoulou G., Papadimitriou K., **Alexandraki V.**, Mavrogonatou E., Alexopoulou K., Anastasiou R., Georgalaki M., Kletsas D., Tsakalidou E., Giaouris E. (2020) The microbiota of Kalathaki and Melichloro Greek artisanal cheeses comprises functional lactic acid bacteria. *LWT Food Sci Technol, 130, 109570*. doi: 10.1016/j.lwt.2020.109570
3. Chatzidaki M., Balkiza F., Gad E., **Alexandraki V.**, Avramiotis S., Georgalaki M., Papadimitriou V., Tsakalidou E., Papadimitriou K., Xenakis A. (2019) Reverse micelles as nano-carriers of nisin against foodborne pathogens. Part II: The case of essential oils. *Food Chem, 28:415-423*. doi: 10.1016/j.foodchem.2018.11.078
4. **Alexandraki V.**, Moatsou G. (2018) Para-κ-casein during the ripening and storage of low-pH, high-moisture Feta cheese. *J Dairy Res, 85(2):226-231*. doi: 10.1017/S0022029918000237
5. Chatzidaki M., Papadimitriou K., **Alexandraki V.**, Balkiza F., Georgalaki M., Papadimitriou V., Tsakalidou E., Xenakis A. Reverse micelles as nanocarriers of nisin against foodborne pathogens (2018) *Food Chem, 255:9-103*. doi: 10.1016/j.foodchem.2018.02.053
6. Kazou M., **Alexandraki V.**, Blom J., Pot B., Tsakalidou E., Papadimitriou K. (2018) Comparative genomics of *Lactobacillus acidipiscis* ACA-DC 1533 isolated from traditional Greek Kopanisti cheese against species within the *Lactobacillus salivarius* clade. *Front Microbiol, 9:1244*. doi: 10.3389/fmicb.2018.01244

7. Zoumpopoulou G., Tzouvanou A., Mavrogonatou E., **Alexandraki V.**, Georgalaki M., Anastasiou R., Papadelli M., Manolopoulou E., Kazou M., Kletsas D., Papadimitriou K., Tsakalidou E. (2017) Probiotic features of lactic acid bacteria isolated from a diverse pool of traditional Greek dairy products regarding specific strain-host interactions. *Probiotics & Antimicro Prot,* doi: 10.1007/s12602-017-9311-9

8. Georgalaki M., Zoumpopoulou G., Mavrogonatou E., Driessche G., **Alexandraki V.**, Anastasiou R., Papadelli M., Kazou M., Manolopoulou E., Kletsas D., Devreese B., Papadimitriou K., Tsakalidou E. (2017) Evaluation of the antihypertensive angiotensin-converting enzyme inhibitory (ACE-I) activity and other probiotic properties of lactic acid bacteria isolated from traditional Greek dairy products. *Int Dairy J, 75, 10-21*. doi: 10.1016/j.idairyj.2017.07.003

9. Kazou M., **Alexandraki V.**, Pot B., Tsakalidou E., Papadimitriou K. (2017) Complete genome sequence of the sourdough isolate *Lactobacillus zymae* ACA-DC 3411. *Genome Announc, 5(30):e00699-17*. doi: 10.1128/genomeA.00699-17

10. Kazou M., **Alexandraki V.**, Pot B., Tsakalidou E., Papadimitriou K. (2017) Whole-genome sequence of the cheese isolate *Lactobacillus rennini* ACA-DC 565. *Genome Announc, 5(5):e01579-16*. doi: 10.1128/genomeA.01579-16

11. Kazou M., **Alexandraki V.**, Pot B., Tsakalidou E., Papadimitriou K. (2017) Complete genome sequence of the dairy isolate *Lactobacillus acidipiscis* ACA-DC 1533. *Genome Announc, 5(4):e01533-16*. doi: 10.1128/genomeA.01533-16

12. Angelopoulou A., **Alexandraki V.**, Georgalaki M., Anastasiou R., Manolopoulou E., Tsakalidou E., Papadimitriou K. (2017) Production of probiotic Feta cheese using *Propionibacterium freudenreichii* subsp. *shermanii* as adjunct. *Int Dairy J, 66, 135-139*. doi: 10.1016/j.idairyj.2016.11.011

13. Chatzidaki M., Papadimitriou K., **Alexandraki V.**, Tsirvouli E., Chakim Z., Ghazal A., Mortensen K., Yaghmur A., Salentinig S., Papadimitriou V., Tsakalidou E., Xenakis A. (2016) Microemulsions as potential carriers of nisin: effect of composition on the structure and efficacy. *Langmuir, 32(35):8988-98.* doi: 10.1021/acs.langmuir.6b02923

14. Anastasiou R., Van Driessche G., Boutou E., Kazou M., **Alexandraki V.**, Vorgias C.E., Devreese B., Tsakalidou E., Papadimitriou K. (2015) Engineered strains of *Streptococcus macedonicus* towards an osmotic stress resistant phenotype retain their ability to produce the bacteriocin macedocin under hyperosmotic conditions. *J. Biotechnol. 212, 125-133*. doi: 10.1016/j.jbiotec.2015.08.018

15. Papadimitriou K., Zoumpopoulou G., Foligné B., **Alexandraki V.**, Kazou M., Pot B., Tsakalidou E. (2015) Discovering probiotic microorganisms: *in vitro, in vivo* and *omics* approaches. *Front. Microbiol. 6:58.* doi: 10.3389/fmicb.2015.00058 *(invited paper)*

16. Alexandraki V., Georgalaki M., Papadimitriou K., Anastasiou R., Zoumpopoulou G., Chatzipavlidis I., Papadelli M., Vallis N., Moschochoritis K., Tsakalidou E. (2014) Determination of triterpenic acids in natural and alkaline-treated Greek table olives throughout the fermentation process. *LWT Food Sci. Technol. 58, 609-613*. https://doi.org/10.1016/j.lwt.2014.04.005

II. Book chapters

1. Papadimitriou K., Zoumpopoulou G., Georgalaki M., **Alexandraki V.**, Kazou M., Anastasiou R, Tsakalidou E. (2019) Sourdough Bread. *In: Innovations in Traditional Foods, C.M. Galanakis Ed., Woodhead Publishing (Elsevier). ISBN: 978-0-12-814888-4 (online)*
2. Zoumpopoulou G, Kazou M., **Alexandraki V.**, Angelopoulou A., Papadimitriou K., Pot B., Tsakalidou E. (2018) Probiotics and Prebiotics: an Overview on Recent Trends. In: *Probiotics and Prebiotics for Animal Health and Food Safety, pp. 1-34, B. Biavati and D. Di Gioia Eds., Springer Publishing. DOI: 10.1007/978-3-319-71950-4_1*
3. **Alexandraki V.**, Kazou M., Angelopoulou A., Arena M.P., Capozzi V., Russo P., Fiocco D., Spano G., Papadimitriou K., Tsakalidou E. (2016) The Microbiota of non-Cow Milk and Products. In: *Non Cow Milk and Milk Products, pp. 117-160, K. Papadimitriou and E. Tsakalidou Eds., Elsevier, San Diego CA US*
4. **Alexandraki V.**, Tsakalidou E. (2015) Kefalograviera Cheese. In: *The Oxford Companion to Cheese, pp., M. Sinsheimer Ed., Oxford University Press USA, NY. (2017 James Beard Award in Reference and Scholarship)*
5. Papadimitriou K., **Alexandraki V.**, Tsakalidou E. (2013) Fighting off human infections: a new role for bacteriocin molecules. In: *Interactive Probiotics, pp. 22-51, E. Pessione Ed., CRC Press*

III. International conference papers

1. **Alexandraki V.**, Kazou M., Charmpi C., Bounenni R.-E., Georgalaki M., Anastasiou R., Tsakalidou E., Papadimitriou K. (2017) Culture independent omics analysis of Feta cheese microbial ecosystem. *4th International Conference on Microbial Diversity, 24-26 October, Bari, Italy* (poster)
2. Kazou M., **Alexandraki V.**, Chavella G., Koutsoumpou M., Paximadaki A., Anastasiou R., Georgalaki M., Tsakalidou E., Papadimitriou K. (2017) Culture-dependent vs metagenomics approaches to discover the table olives microbial ecosystem. *4th International Conference on Microbial Diversity, 24-26 October, Bari, Italy* (poster)
3. Chatzidaki M.D., Papadimitriou K., **Alexandraki V.**, Georgalaki M., Balkiza F., Papadimitriou V., Xenakis A. and Tsakalidou E. (2017) Microemulsions as potential carriers of nisin: effect of composition on structure and efficacy. *10th NIZO Dairy Conference, 1-3 October, Papendal, The Netherlands* (poster)
4. Papadimitriou K., **Alexandraki V.**, Kazou M., Charmpi C., Bounenni R.-E., Georgalaki M., Anastasiou R., Tsakalidou E. (2017) Feta cheese microbial ecosystem: a metagenomics approach. *12th Symposium on Lactic Acid Bacteria, 27-31 August, Egmond aan Zee, The Netherlands* (poster)
5. Papadimitriou K., Kazou M., **Alexandraki V.**, Chavella G., Koutsoumpou M., Paximadaki A., Anastasiou R., Georgalaki M., Tsakalidou E. (2017) Metagenomics analysis of naturally fermented Greek table olives. *12th Symposium on Lactic Acid Bacteria, 27-31 August, Egmond aan Zee, The Netherlands* (poster)

6. Papadimitriou K., Chatzidaki D.M., **Alexandraki V.**, Georgalaki M., Balkiza F., Papadimitriou V., Xenakis A., Tsakalidou E. (2017) Structure and antimicrobial efficacy of microemulsions carrying nisin. *12th Symposium on Lactic Acid Bacteria, 27-31 August, Egmond aan Zee, The Netherlands* (poster)

7. Papadimitriou K., Kazou M., **Alexandraki V.**, Pot B., Tsakalidou E. (2016) *In silico* analysis of the first complete genome sequence of *Lactobacillus acidipiscis* species". *41st Congress of Federation of European Biochemical Societies (FEBS), 3-8 September, Kusadasi, Turkey* (poster)

8. Ioannou M., Zoumpopoulou G., Moschopoulou E., Anastasiou R., **Alexandraki V.**, Tsakalidou E., Papadimitriou K. (2016) Kaimaki type ice cream as a food carrier of the probiotic strain *Lactobacillus fermentum* ACA-DC 179. *12th International Association for Food Protection (IAFP), 11-13 May 2016, Athens, Greece* (poster)

9. **Alexandraki V.**, Kazou M., Charmpi C., Bounenni R.-E., Georgalaki M., Anastasiou R., Tsakalidou E., Papadimitriou K. (2016) Metagenomics analysis of the Feta cheese microbial ecosystem. *12th International Association for Food Protection (IAFP), 11-13 May 2016. Athens, Greece* (poster)

10. Kazou M., **Alexandraki V.**, Pot B., Tsakalidou E., Papadimitriou K. (2016) Complete genome sequence of the dairy isolate *Lactobacillus acidipiscis* ACA-DC 1533. *12th International Association for Food Protection (IAFP), 11-13 May 2016, Athens, Greece* (poster)

11. Kazou M., **Alexandraki V.**, Koutsoumpou M., Paximadaki A., Anastasiou R., Georgalaki M., Tsakalidou E., Papadimitriou K. (2016) Exploring the microbial consortia of Greek table olives using culture-dependent and -independent approaches. *12th International Association for Food Protection (IAFP), 11-13 May 2016, Athens, Greece* (poster)

12. Georgalaki M., Zoumpopoulou G., Mavrogonatou E., Van Driessche G., Anastasiou R., **Alexandraki V.**, Kazou M., Papadelli M., Manolopoulou E., Papadimitriou K., Kletsas D., Devreese B., Tsakalidou E. (2016) Production of bioactive peptides and probiotic potential of lactic acid bacteria isolated from traditional Greek dairy products. *12th International Association for Food Protection (IAFP), 11-13 May 2016, Athens, Greece* (poster)

13. Papadimitriou K., Kazou M., **Alexandraki V.**, Pot B., Renault P., Tsakalidou E. (2015) Genomics of *Streptococcus macedonicus*: moving from pathogenicity to adaptation to the dairy environment. *FEMS 6th Congress of European Microbiologists, 7-11 June, Maastricht, the Netherlands* (poster)

14. Zoumpopoulou G., **Alexandraki V.**, Kazou M., Papadelli M., Tzouvanou A., Manolopoulou E., Anastasiou R., Georgalaki M., Mavrogonatou E., Kletsas D., Papadimitriou K., Tsakalidou E. (2015) Greek traditional dairy and meat products: A biological reservoir for new probiotic strains. *FEMS 6th Congress of European Microbiologists, 7-11 June, Maastricht, the Netherlands* (poster)

15. Georgalaki M., Kazou M., **Alexandraki V.**, Manolopoulou E., Anastasiou R., Zoumpopoulou G., Papadimitriou K., Papadelli M., Van Driessche G., Devreese B., Tsakalidou E. (2015) Production of ACE-inhibitory peptides by lactic acid bacteria isolated from traditional Greek

dairy products. *FEMS 6th Congress of European Microbiologists, 7-11 June, Maastricht, the Netherlands* (poster)

16. Zoumpopoulou G., Papadelli M., Tzouvanou A., **Alexandraki V.**, Kazou M., Manolopoulou E., Anastasiou R., Georgalaki M., Papadimitriou K., Tsakalidou E. (2015) *In vitro* screening for probiotic potential of lactic acid bacteria isolated from Greek traditional dairy and meat products. *7th IDF International Symposium on Sheep, Goat and other non-Cow Milk, 23-25 March, Limassol, Cyprus* (poster)
17. Georgalaki M., Manolopoulou E., Anastasiou R., Zoumpopoulou G., **Alexandraki V.**, Kazou M., Papadimitriou K., Papadelli M., Van Driessche G., Devreese B., Tsakalidou E. (2015) Production of ACE-Inhibitory peptides by lactic acid bacteria isolated from Greek traditional yogurt and fermented milk samples. *7th IDF International Symposium on Sheep, Goat and other non-Cow Milk, 23-25 March, Limassol, Cyprus* (poster)
18. Georgalaki M., Manolopoulou E., Anastasiou R., Zoumpopoulou G., **Alexandraki V.**, Kazou M., Papadimitriou K., Papadelli M., Tsakalidou E. (2014) ACE-Inhibitory activity and technological potential of lactic acid bacteria isolated from Greek traditional yogurt and fermented milk samples. *11th Symposium on Lactic Acid Bacteria, 31 August - 4 September, Egmond aan Zee, The Netherlands* (poster)
19. Zoumpopoulou G., Papadelli M., Tzouvanou A., **Alexandraki V.**, Kazou M., Manolopoulou E., R. Anastasiou, Georgalaki M., Papadimitriou K., Tsakalidou E. (2014) Probiotic traits of lactic acid bacteria isolated from Greek traditional dairy and meat products. *International Scientific Conference on Probiotics and Prebiotics - IPC2014, 24 - 26 June, Budapest, Hungary* (poster)
20. Anastasiou R., Kazou M., **Alexandraki V.**, Tsakalidou E., Papadimitriou K. (2014) Linking stress resistant phenotypes to bacteriocin production under stress conditions: the example of *Streptococcus macedonicus* osmotic stress resistant mutants and macedocin. *4th International Symposium on Antimicrobial Peptides, 4-6 June, Lorient, France* (poster)
21. Georgalaki M., Zoumpopoulou G., Papadimitriou K., **Alexandraki V.**, Kazou M., Van Driessche G., Devreese B., Tsakalidou E. (2014). Is the antimicrobial activity exhibited by *Lactobacillus fermentum* ACA-DC 179 attributed to a bacteriocin? *4th International Symposium on Antimicrobial Peptides, 4-6 June, Lorient, France* (poster)
22. **Alexandraki V.**, Moatsou G. (2011) Detection of caprine milk in ovine milk. *IDF International Symposium on Sheep, Goat and other non-Cow Milk, 16-18 May, Athens, Greece* (poster)

IV. National conference papers

1. Papadimitriou K., Georgalaki M., Zoumpopoulou G., Mavrogonatou E., Driessche G., **Alexandraki V.**, Anastasiou R., Papadelli M., Kazou M., Manolopoulou E., Kletsas D., Devreese B., Tsakalidou E. (2016) Evaluation of angiotensin-converting enzyme-inhibitory (ACE-I) activity and other probiotic properties of lactic acid bacteria isolated from traditional Greek dairy products. *67th Conference of Hellenic Society for Biochemistry and Molecular Biology (HSBMB), 25-27 November, Ioannina, Greece* (poster)

2. Charmpi C., Anastasiou R., Georgalaki M., Kazou M., **Alexandraki V.**, Papadimitriou K., Tsakalidou E. (2015) The microbiota of Greek Feta cheese - Classical and metagenomics approaches. *66th Conference of Hellenic Society for Biochemistry and Molecular Biology (HSBMB), 11-13 December, Athens, Greece* (poster)

3. Bounenni R.-E., Anastasiou R., Georgalaki M., Kazou M., **Alexandraki V.**, Papadimitriou K., Tsakalidou E. (2015) Classical versus metagenomics analysis of the microbiota of traditional Feta cheese. *66th Conference of Hellenic Society for Biochemistry and Molecular Biology (HSBMB), 11-13 December, Athens, Greece* (poster)

4. Koutsoumpou M., Anastasiou R., Georgalaki M., Kazou M., **Alexandraki V.**, Papadimitriou K., Tsakalidou E. (2015) Exploring the microbial ecosystem of naturally fermented Greek table olives. *66th Conference of Hellenic Society for Biochemistry and Molecular Biology (HSBMB), 11-13 December, Athens, Greece* (poster)

5. Kazou M., **Alexandraki V.**, Papadimitriou K., Tsakalidou E. (2015) Exploring the olive ecosystem. The transition from classical microbiological analysis to metagenomics analysis. *Workshop on "Bioactive compounds from olive and olive oil" 31 October, Monemvasia, Greece* (oral presentation)

6. Mylona M., Alexopoulou A., **Alexandraki V.**, Tsakalidou E., Papadimitriou K., Giaouris E. (2015) Study of endogenous microflora of traditional cheeses kalathaki and melichloro, isolation of wild strains of lactic acid bacteria and identification of them by molecular techniques. *6th National MicroBioKosmos (MBK) Conference, 3-5 April, Athens, Greece* (poster)

7. Kazou M., **Alexandraki V.**, Sarafianou A., Pot B., Renault P., Tsakalidou E., K. Papadimitriou (2014) Comparative genomic analysis among three dairy *Streptococcus macedonicus* strains. *65th Conference of Hellenic Society for Biochemistry and Molecular Biology, 28-30 November, Thessaloniki, Greece* (poster)

8. Zoumpopoulou G., **Alexandraki V.**, Kazou M., Papadelli M., Tzouvanou A., Manolopoulou E., Anastasiou R., Georgalaki M., Mavrogonatoy E., Kletsas D., Papadimitriou K., Tsakalidou E. (2014) *In vitro* evaluation of probiotic attributes of lactic acid bacteria isolated from Greek traditional dairy and meat products. *65th Conference of Hellenic Society for Biochemistry and Molecular Biology, 28-30 November, Thessaloniki, Greece* (poster)

9. Kazou M., Peton V., Jardin J., Gagnaire V., **Alexandraki V.**, Tsakalidou E., Jan G., Papadimitriou K. (2013) Proteome changes of *Lactococcus lactis* MG1363 during exposure to the bacteriocin peptide macedocin produced by *Streptococcus macedonicus* ACA-DC 198. *64th Conference of Hellenic Society for Biochemistry and Molecular Biology (HSBMB), 6-8 December, Athens, Greece* (poster)

10. **Kazou M.,** Zoumpopoulou G., Alexandraki V., Tarantilis P.A., Polissiou M., Tsakalidou E., Papadimitriou K. (2013) Global cellular changes of *Lactococcus lactis* subsp. *cremoris* MG1363 during exposure to different stress stimuli assessed by FT-IR spectroscopy. *8th Meeting of the Hellenic Society for Computational Biology and Bioinformatics (HSCBB), 22-24 November, Lamia, Greece* (oral presentation)

11. **Alexandraki V.**, Georgalaki M., Papadimitriou K., Anastasiou R., Zoumpopoulou G., Chatzipavlidis I., Papadelli M., Vallis N., Moschochoritis K., Tsakalidou E. (2013) Does the treatment affect the triterpenic acid content in table olives? *OLITEC | International Workshop in Bioactive compounds from Olea europea: Chemistry and Biology, 12-13 May, Athens, Greece* (poster)
12. Georgalaki M., **Alexandraki V.**, Papadimitriou K., Anastasiou R., Zoumpopoulou G., Chatzipavlidis I., Papadelli M., Vallis N., Moschochoritis K., Tsakalidou E. (2011) Detection of Triterpenic Acids in Greek Table Olives. *International Year of Chemistry-Days of Chemistry 2011, Athens, Greece* (oral presentation)

V. Other publications

1. Papadimitriou K., Xatzidaki M., **Alexandraki V.**, Georgalaki M., Papadimitriou V., Tsakalidou E., Xenakis A., "Water-in-oil microemulsions as bacteriocins' carriers for use as "dressing" type products with antimicrobial food protection", *Patent 1008858, issued September 27, 2016*
2. Abeliotis K., Vallis N., Moschochoritis K., Tsakalidou E., **Alexandraki V.** (2014) Product Category Rules according to ISO 14025:2006, *Product Group: UN CPC 21340 & 21399 Table Olives* (expired 2016-06-13; being updated)
3. **Alexandraki V.**, Tsakalidou E., Papadimitriou K., Holzapfel W. (2013) A study on the status and trends of the conservation and sustainable use of microorganisms in food processes. *Background Study Paper N^{o} 65, FAO Commission on Genetic Resources for Food and Agriculture* (invited paper)

Contents

List of abbreviations

ABC	ATP-binding cassette
ABySS	assembly by short sequences
ADI	arginine deiminase
Ala	alanine
ANI	average nucleotide identity
ArAT	aromatic aminotransferase
Arg	arginine
Asn	asparagine
Asp	aspartate
AspAT	aspartate aminotransferase
BCAA	branched-chain amino acid
BcAT	branched-chain aminotransferase
BLASR	basic local alignment with successive refinement
BLAST	basic local alignment search tool
BLASTP	Protein BLAST
BPGA	bacterial pan genome analysis
CAGR	compound annual growth rate
Cas	CRISPR associated
CDD	conserved domain database
CDS	coding sequence
CEP	cell-envelope proteinase
CK	carbamate kinase
CLR	continuous long read
COG	clusters of orthologous groups
CPS	capsular polysaccharides
CRISPR	clustered regularly interspaced short palindromic repeat
crRNA	CRISPR RNA
DELTA-BLAST	domain enhanced lookup time accelerated BLAST
DFM	direct-fed microbials
dNTP	deoxynucleotide
DO	dissolved oxygen
EFFCA	European Food and Feed Cultures Association
EFSA	European Food Safety Authority
EMP	Embden-Meyerhof-Parnas

EPS	exopolysaccharide
FAA	free amino acid
FAO	Food and Agriculture Organization
FDA	Food and Drug Administration
FFA	free fatty acid
GABA	γ-aminobutyrate
GAD	γ-aminobutyrate decarboxylase
GDH	glutamate dehydrogenase
GI	genomic island
GIT	gastrointestinal tract
Gln	glutamine
Glu	glutamate
Gly	glycine
GO	gene ontology
GOLD	genomes online database
GRAS	generally recognized as safe
GSH	glutathione
HGT	horizontal gene transfer
HDC	histidine decarboxylase
His	histidine
HTS	high-throughput sequencing
IMG	integrated microbial genomes
KEGG	Kyoto encyclopedia of genes and genomes
KO	KEGG orthology
KOALA	KEGG orthology and links annotation
LAB	lactic acid bacteria
Lys	lysine
MBGD	microbial genome database
MDR	multidrug-resistant
MiGA	microbial genomes atlas
MLST	multilocus sequence typing
NCBI	National Center for Biotechnology Information
NGS	next-generation sequencing
nr	non-redundant
NSLAB	non-starter LAB
ONT	Oxford Nanopore Technologies

ORF	open reading frame
OTC	ornithine carbamoyl-transferase
PacBio	Pacific Biosciences
PDB	protein data bank
PEP/PTS	phosphoenolpyruvate/phosphotransferase
Pfam	protein family database
PFL	pyruvate formate lyase
PFLA	pyruvate formate-lyase activating
PGAP	prokaryotic genome annotation pipeline
PHASTER	PHAge Search Tool-Enhanced Release
PHI-BLAST	Pattern-Hit Initiated BLAST
PHYLIP	phylogeny inference package
PMF	proton motive force
PP	pentose-phosphate
pre-crRNA	precursor CRISPR RNA
Pro	proline
PSI-BLAST	position-specific iterative BLAST
QPS	qualified presumption of safety
R-M	restriction-modification
RedSeq	reference sequence database
ROS	reactive oxygen species
rRNA	ribosomal RNA
tRNA	transfer RNA
Ser	serine
SHMT	serine hydroxymethyltransferase
SLAB	starter LAB
SMRT	single molecule real time
SOAP	short oligonucleotide analysis package
SOLiD	sequencing by oligo ligation and detection
Thr	threonine
UniProtKB	universal protein resource knowledgebase
WGM	whole genome mapping
WGS	whole-genome sequencing
WHO	World Health Organization
ZMW	zeromode waveguide

1

Introduction and aim of the thesis

1.1 Introduction

1.1.1 Lactic acid bacteria

Lactic acid bacteria (LAB) comprise an ancient heterogeneous group of microorganisms, which share a number of morphological and physiological traits. According to the current taxonomy, they belong to the phylum *Firmicutes*, class *Bacilli*, and order *Lactobacillales*. The different families include the genera *Aerococcus, Carnobacterium, Enterococcus, Lactobacillus, Lactococcus, Leuconostoc, Oenococcus, Pediococcus, Streptococcus, Tetragenococcus, Vagococcus* and *Weissella* (von Wright and Axelsson, 2019). Although the genus *Bifidobacterium* is actually a member of the phylum *Actinobacteria* and is phylogenetically distinct from LAB, belonging to the high G+C content (ranging from 42 to 67%) Gram-positive bacteria (Biavati, 2001), it is often included in this group, due to the production of lactic acid as one of the main fermentation end products (Pokusaeva et al., 2011). The phylogenetic relationship among some of the main genera in the order *Lactobacillales* is illustrated in Figure 1.

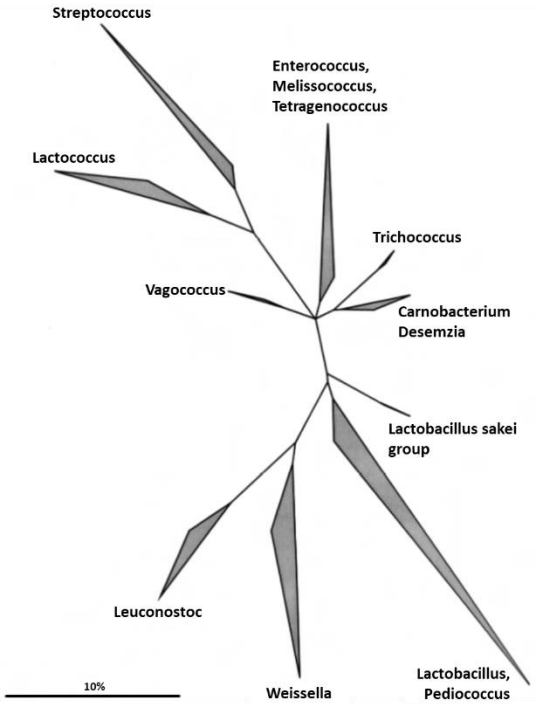


Figure 1. Phylogenetic tree depicting the position of LAB genera in relation to each other (Hammes and Hertel, 2006)

LAB are gram-positive, catalase-negative, mostly non-motile, non-sporulating, anaerobic or aerotolerant cocci or rods. They lack the ability to synthesize components of the respiratory chains, such as porphyrin and cytochromes, and therefore cannot generate ATP by creation of a proton gradient (von Wright and Axelsson, 2019). They occupy various ecological niches and their key characteristic is the ability to ferment carbohydrates to lactic acid as the main product of sugar metabolism (George et al., 2018; von Wright and Axelsson, 2019). Several metabolic processes carried out during their growth (e.g. rapid acidification due to lactose fermentation, proteolysis, amino acid catabolism, exopolysaccharide production) affect several quality related aspects of the produced fermented food, such as flavor, texture and nutritional value (Nout, 2009; Liu et al., 2011; Kongo, 2013). Thus, they play a significant role in numerous industrial food fermentations where they are used as starter and adjunct cultures (Leroy and De Vuyst, 2004; Divya et al., 2012; Gobbetti et al., 2015; Kongo, 2013).

Furthermore, the indigenous LAB microbiota is responsible for the spontaneous fermentation of raw food materials, which results in the production of several traditional fermented products worldwide. These products constitute diverse biological pools for the isolation on new LAB strains with potent technological and probiotic properties that could be used as starters in commercial fermentations (Cogan et al., 1997; De Vuyst et al., 2002; Liu et al., 2011; Ramraj et al., 2012; Mokoena et al., 2016; Zoumpopoulou et al., 2018b). LAB species commonly identified in fermented foods are presented in Table 1.

1.1.2 Metabolism of lactic acid bacteria

1.1.2.1 Carbohydrate fermentation

LAB acquire energy from substrate level phosphorylation (i.e. fermentation). The sugars are either accumulated intracellularly and then phosphorylated by ATP-dependent kinases or they are phosphorylated during translocation via phosphoenolpyruvate/phosphotransferase (PEP/PTS) systems (Thompson, 1987). They utilize two basic fermentative pathways, the homofermentative, which is based on glycolysis (Embden-Meyerhof-Parnas pathway, EMP) and produces only lactic acid, and the heterofermentative one (pentose-phosphate pathway, PP) with lactic acid, CO_2 and ethanol or acetic acid as end products (Vijayakumar et al., 2008). Other

hexoses than glucose, such as mannose and fructose, enter these pathways after different isomerization and phosphorylation steps as either glucose-6-phosphate or fructose-6-phosphate. In contrast to hexoses, which can be fermented via both pathways, pentoses can only be fermented heterofermentatively by entering the pathway as ribulose-5-phosphate or xylulose-5-phosphate, with no CO_2 production (Von Wright and Axelsson, 2019).

Table 1. LAB species in common fermented foods (Divya et al., 2012)

Fermented foods	Reported LAB strains
Yogurt	*Lb. delbrueckii* subsp*. bulgaricus, S. thermophilus*
Cheddar cheese	*L. lactis* subsp*. lactis, L. lactis* subsp*. cremoris, S. thermophilus*
Italian cheese (Mozzarella)	*Lb. delbrueckii* subsp*. bulgaricus, Lb. helveticus, Lb. lactis, S. thermophilus*
Swiss cheese types	*Lb. delbrueckii* subsp*. bulgaricus, Lb. lactis, L. lactis* subsp*. biovar diacetylactis, Lc. mesenteroides* subsp*. cremoris, L. lactis* subsp*. lactis, L. lactis* subsp*. cremoris, S. thermophilus*
Goat cheese and sheep cheese	*L. lactis* subsp*. lactis, L. lactis* subsp*. cremoris, L. lactis* subsp*. biovar diacetylactis, Lc. mesenteroides* subsp*. cremoris*
Butter and buttermilk	*L. lactis* subsp*. lactis, L. lactis* subsp*. lactis biovar diacetylactis, L. lactis* subsp*. cremoris, Lc. menesteroides* subsp*. cremoris*
Kefir	*Lb. kefir, Lb. kefiranofacies, Lb. brevis*
Fermented, probiotic milk	*Lb. casei, Lb. acidophilus, Lb. rhamnosus, Lb. johnsonii*
Fermented sausages	*Lb. sakei, Lb. curvatus, P. acidilactici, P. pentosaceus*
Sauerkraut	*Lb. plantarum, Lb. brevis, P. acidilactici, P. cerevisiae, Lc. mesenteroides*
Pickles	*Lb. plantarum, Lb. pentosus, Lb. plantarum, Lc. mesenteroides*
Kimchi (Korea)	*Lb. plantarum, Lc. mesenteroides, Lb. brevis*
Idli/dosa (India)	*Lc. mesenteroides, E. faecalis*
Wine	*O. oeni*
Rice wine	*Lb. sakei*

E.: *Enterococcus*; **L.:** *Lactococcus*; **Lb.:** *Lactobacillus*; **Lc.:** *Leuconostoc*, **O.:** *Oenococcus*, **P.:** *Pediococcus*, **S.:** *Streptococcus*

Obligatory homofermentative LAB include, among others, *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus, Lactobacillus acidophilus*, and *Lactobacillus helveticus*. These species cannot metabolize pentoses. Obligatory heterofermentative LAB include the genera *Leuconostoc*, *Oenococcus*, *Weissella*, and *Fructobacillus* as well as the species *Lactobacillus brevis*, *Lactobacillus fermentum* and *Lactobacillus reuteri*. Finally, most of other LAB including *Lactobacillus casei*, *Lactobacillus curvatus*, *Lactobacillus plantarum*, *Lactobacillus rhamnosus*, *Lactobacillus sakei*, *Lactococcus lactis*, *Lactobacillus pentosus* and *Lactobacillus xylosus* are facultatively heterofermentative since they ferment hexoses homofermentatively,

but may also ferment pentoses through heterolactic fermentation (Hammes and Vogel, 1995; von Wright and Axelsson, 2019).

The uptake of lactose by the bacterial cells is facilitated either by a lactose-specific phosphoenolpyruvate/phosphotransferase (PEP/PTS) system (*lacFE*) or by a non-specific (*galP*) or lactose-specific (*lacS*, found in *S. thermophilus*) permeases (Figure 2).

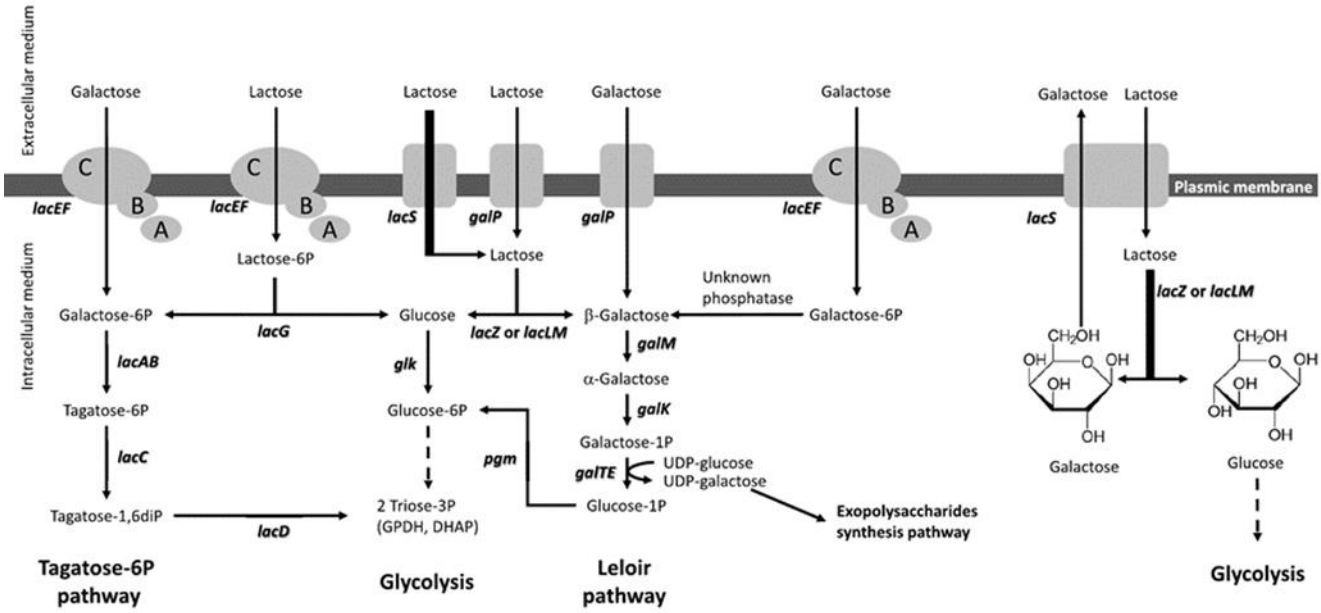


Figure 2. Schematic representation of lactose and galactose catabolic pathways (Iskandar et al., 2019)

When lactose is transferred via the PEP/PTS system (*lacFE*), as lactose phosphate, it is hydrolyzed by phospho-β-galactosidase (*lacG*) into glucose and galactose-6-phosphate. Glucose is then phosphorylated by glucokinase (*glk*) and enters either the EMP or PP pathway, depending on the fermentation mechanism employed by LAB, while galactose-6-phosphate is metabolized through the tagatose phosphate pathway. More specifically, galactose-6-phosphate is transformed into the trioses glyceraldehyde-3-phosphate and dihydroxyacetone phosphate via a series of reactions catalyzed by galactose-6-phosphate isomerase (*lacAB*), tagatose-6-phosphate kinase (*lacC*) and tagatose-1,6-diphosphate aldolase (*lacD*). Free galactose can also be transferred via the PEP/PTS system, as galactose phosphate, and metabolized either through the tagatose or the Leloir pathway (Figure 2; Kandler, 1983; Iskandar et al., 2019; von Wright and Axelsson, 2019).

If the transport of lactose is permease mediated (*galP* or *lacS*), the sugar is cleaved by β-galactosidase (*lacZ* or *lacLM*) to glucose and β-galactose. The former enters the fermentation pathways mentioned above, while the latter is metabolized through the Leloir pathway, encoded by *galMKTE* (Figure 2). More specifically, galactose is metabolized into glucose-1-phosphate with the enzymes galactose mutarotase (*galM*), galactose kinase (*galK*) and galactose-1-phosphate uridyltransferase/UDP-glucose-4-epimerase (*galT/galE*). Finally, glucose-1-phosphate is then transformed into glucose-6-phosphate by the action of the phosphoglucomutase (*pgm*). The majority of *S. thermophilus* strains and thermophilic lactobacilli, e.g. *Lb. bulgaricus*, ferment lactose partially, since galactose is not metabolized but released and accumulated into the medium (Iskandar et al., 2019). However, it has been reported that Gal^+ *S. thermophilus* strains can also metabolize the galactose moiety (de Vin et al., 2005).

Although the majority of LAB acquire energy solely from fermentation metabolism, many species are genetically equipped (*cydABCD* genes) for respiration metabolism (Pedersen et al., 2012). The latter is activated in the presence of either exogenous heme (e.g. *L. lactis*, *Enterococcus faecalis*, *Leuconostoc mesenteroides*, *Weissella cibaria*) or heme and menaquinone (e.g. *Lb. brevis, Lb. casei, Lb. fermentum, Lactobacilus paracasei, Lb. plantarum*; Pedersen et al., 2012).

1.1.2.2 Citrate metabolism

Apart from the production of lactate, the technological importance of certain LAB, like lactococci and leuconostocs, is also associated with the production of diacetyl, acetaldehyde, acetoin and 2,3-butanediol, which are the main flavor and aroma compounds in dairy products, through citrate metabolism (Urbach, 1995; García Quintáns et al., 2008). Although citrate is present in milk in relatively low concentration (∼8 mmol L^{-1}) and approximately 90% is lost in the whey during cheese making, its concentration in the aqueous phase of cheese, presumably as colloidal citrate, is ∼3 times more than that in the whey (McSweeney et al., 2017). Citrate uptake by the bacterial cells is facilitated by a citrate permease in exchange to lactate produced by glycolysis. This efflux of lactate is probably a resistance mechanism for citrate-fermenting LAB under acidic conditions. For example, in *L. lactis* citrate metabolism is induced due to the acidic

conditions of the late exponential growth phase to render cells more resilient to the accumulation of lactate (Magni et al., 1999). In the cell citrate is cleaved into acetate and oxaloacetate by a citrate lyase, and subsequently oxaloacetate is decarboxylated into pyruvate and CO_2 by an oxaloacetate decarboxylase. The metabolism of pyruvate results in the production of aroma and flavor compounds (Drider et al, 2004; García Quintáns et al., 2008), as shown in Figure 3.

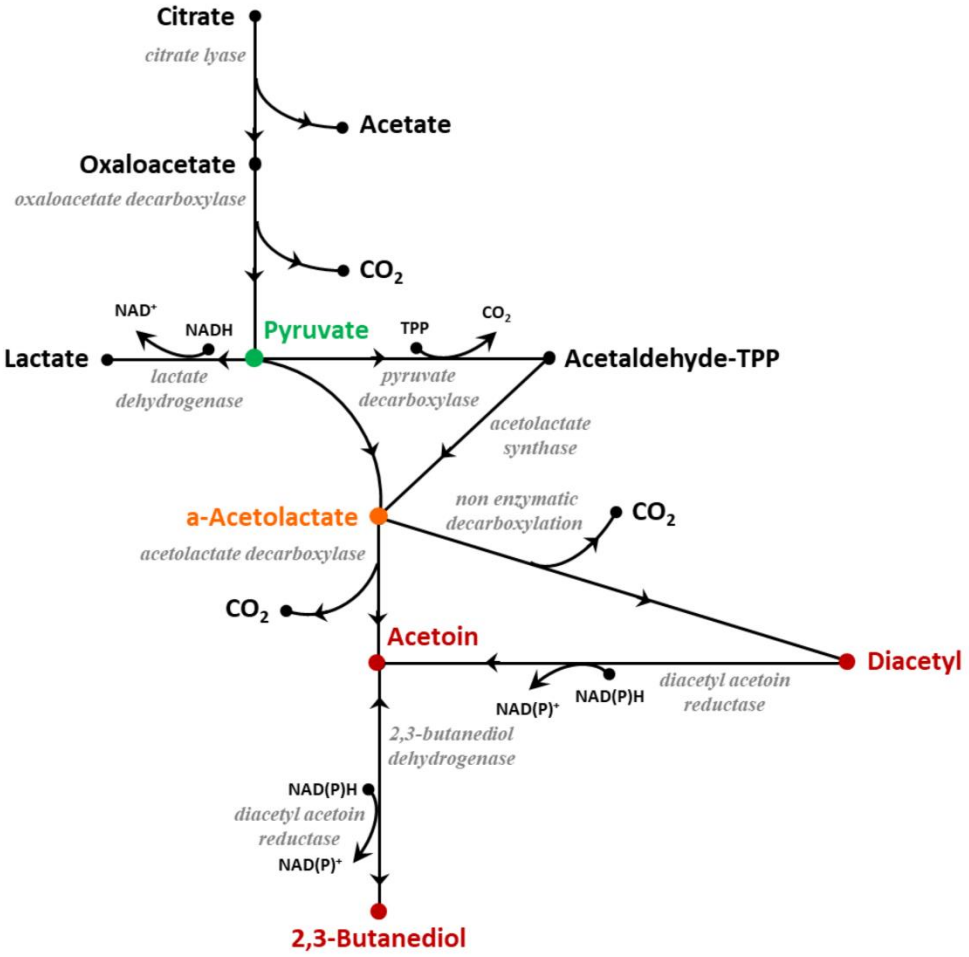


Figure 3. Production of aroma and flavor compounds from citrate metabolism in LAB (García Quintáns et al., 2008)

Strains of *L. lactis* subsp. *lactis biovar diacetylactis*, and some species of the genera *Leuconostoc* and *Weissella* are used as diacetyl producers in the dairy industry (García Quintáns et al., 2008). In addition, it has been demonstrated that *Enterococcus* species like *E. faecalis* and

Enterococcus faecium are able to metabolize citrate and therefore to actively contribute to the aroma and flavor development of fermented dairy products (Sarantinopoulos et al., 2001; Martino et al., 2016).

In contrast to dairy products, citrate fermentation can adversely affect flavor, as in the case of the malolactic bacterium *Oenococcus oeni*. The consumption of citrate changes the aroma profile of wines due to the production of volatile compounds, such as diacetyl, which confers an undesirable ‘‘buttery’’ character to wine (Olguin et al., 2009).

1.1.2.3 Proteolytic system

The proteolytic system of LAB utilizes proteins and peptides and supplies the bacterial cells with essential amino acids, which facilitate bacterial growth (Savijoki et al., 2006; Liu et al., 2010). Furthermore, many flavor and aroma compounds are produced from amino acids catabolism, which shape the organoleptic characteristics of the fermented products (Thierry et al., 2015).

The proteolytic system of dairy LAB is properly equipped for growth in milk and the requirements in amino acids and peptides is met by the proteolysis of caseins, the most abundant protein in milk (Figure 4). Proteolytic cascade is initiated by a cell-envelope proteinase (CEP) that degrades caseins into oligopeptides, di- and tripeptides and amino acids. These enzymes differentiate among LAB species and the characterized types include PrtP (in *L. lactis* and *Lb. paracasei*), PrtH (in *Lb. helveticus*), PrtR (in *Lb. rhamnosus*), PrtS (in *S. thermophilus*), PrtB (in *Lb. bulgaricus*; Savijoki et al., 2006) and PrtL (*Lb. lactis*; Villegas et al., 2015). It should be stressed that the presence of the CEPs in certain LAB species, like *S. thermophilus*, is a strain-depended trait (Liu et al., 2010; Goh et al., 2011).

The oligopeptides produced by CEP are then transferred via specific oligopeptide transport systems (Opp) into the cells. The proteins of these systems belong to the ATP-binding cassette (ABC) transporters superfamily and consist of one extracellular substrate-binding protein (OppA), which binds the oligopeptide, two integral membranes proteins (OppB, OppC) that form the pore and two membrane-bound cytoplasmic ATP-binding proteins (OppD, OppF) that provide the energy for peptide translocation. The genes of the Opp system have been

identified in *Lb. acidophilus*, *Lb. bulgaricus*, *Lb. brevis*, *Lb. casei*, *Lb. rhamnosus*, *Lactobacillus johnsonii*, *Lactobacillus gasseri*, *Lb. helveticus*, *O. oeni* and *L. lactis* subsp. *lactis* (Liu et al., 2010). Moreover, several studies in various LAB species, like *L. lactis*, *Lb. bulgaricus* and *S. thermophilus*, revealed that genes encoding these proteins are organized in an operon (Tynkkynen et al., 1993; Peltoniemi et al., 2002; Gardan et al., 2009).

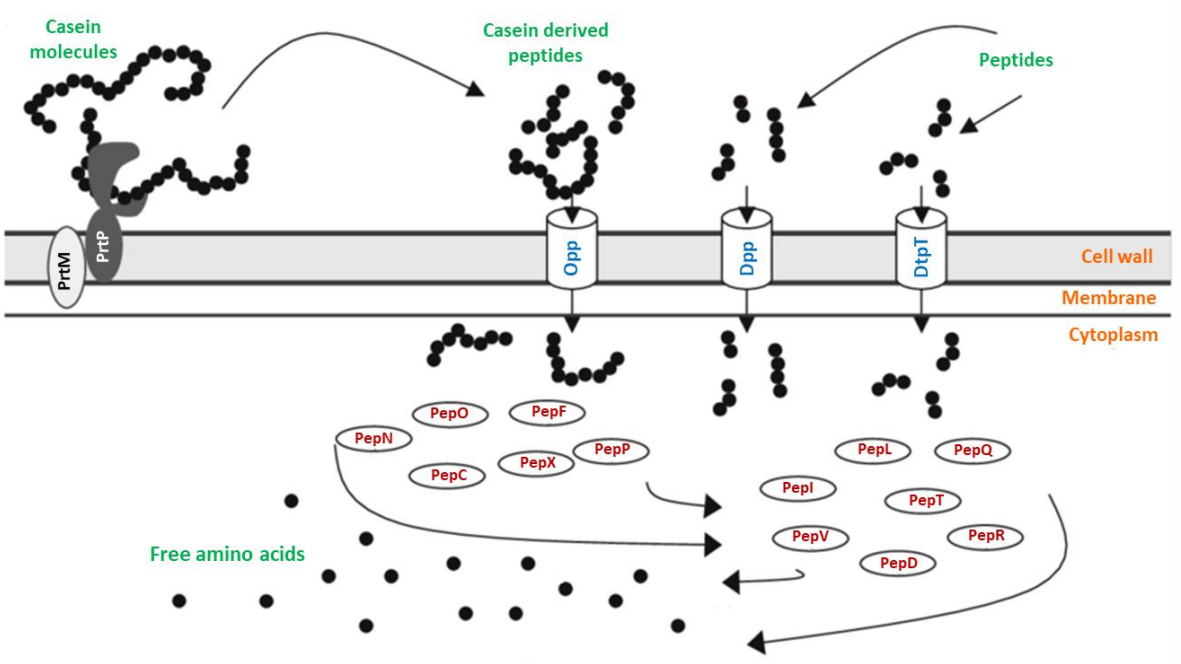


Figure 4. Simplified presentation of the function and regulation of the proteolytic system of lactococci in casein breakdown (Savijoki et al., 2006)

In addition to Opp transport systems, several LAB species also carry transporters, which facilitate the uptake of oligopeptides, such as the Dpp and the DtpT transport systems (Liu et al., 2010; Griffiths and Tellez, 2013; Kazou et al., 2018). The Dpp system is an ATP-driven ABC transporter, belonging to the ABC superfamily. The system is organized similarly to the Opp system and comprises the DppA, which is responsible for binding the di- and tripeptides, the DppB and DppC permeases, and finally, the DppD and DppF ATP-binding proteins (Sanz et al., 2001; Liu et al., 2010). Dpp is capable of transporting di-, tri-, and tetrapeptides, which contain relatively hydrophobic branched-chain amino acids (BCAAs) and displays the highest preference to tripeptides (Savijoki et al., 2006). The Dpp system is widely distributed among LAB species (e.g.

Lb. acidophilus, *Lb. bulgaricus*, *Lb. helveticus*, *Lb. plantarum*, *Lb. brevis*, *Lb. sakei*, *Lactobacillus salivarius*, *Lb. casei*, *Lb. rhamnosus*, *Pediococcus pentosaceus*, *O. oeni*, *Lc. mesenteroides*, *L. lactis* subsp. *lactis*, *L. lactis* subsp. *cremoris*), with the DppA gene being often present in multiple copies. Of note, the respective transporter has not been found in *Lb. johnsonii*, *Lb. gasseri*, *Lb. reuteri*, while in *S. thermophilus* the majority of the Dpp-associated genes are putative pseudogenes (Liu et al., 2010).

DtpT is a proton motive force (PMF)-driven peptide transport system with high affinity for hydrophilic and charged di- and tripeptides (Savijoki et al., 2006). In *L. lactis* the respective gene is located adjacent to the aminopeptidase PepN encoding gene (Hagting et al., 1994). This system has been found in several LAB species, including *Lb. acidophilus*, *Lb. johnsonii*, *Lb. gasseri*, *Lb. plantarum*, *Lb. brevis*, *Lb. reuteri*, *Lb. sakei*, *Lb. salivarius*, *Lb. casei*, *Lb. rhamnosus*, *P. pentosaceus*, *L. lactis* subsp. *lactis*, *L. lactis* subsp. *cremoris* and *S. thermophilus*, but not in *Lb. bulgaricus*, *Lb. helveticus*, *O. oeni* and *Lc. mesenteroides* (Liu et al., 2010).

Inside the bacterial cells, the oligopeptides are cleaved by a number of intracellular peptidases and thus shorter peptides along with amino acids are produced (Liu et al., 2010; Savijoki et al., 2006). General aminopeptidases, i.e. metallopeptidase PepN and cysteine peptidase PepC, along with the X-prolyl dipeptidyl aminopeptidase (PepX) are the first enzymes that cleave the available oligopeptides, resulting in the production of di-and tripeptides. Activity of general aminopeptidases is located on the N-terminal amino acid residues of the peptides and their specificity depends on the peptide length and the nature of the amino acid residues. Di/tripeptides are further degraded by the tripeptidase, PepT, and dipeptidases, PepV and PepD. These enzymes prefer peptides that contain hydrophobic amino acids including leucine, methionine, phenylalanine, or glycine. The remaining endopeptidases present high substrate specificity. PepA releases N-terminal acidic residues from peptides (3-9 residues) and PepS prefers smaller peptides (2-5 residues) with arginine (Arg) or aromatic amino acid residues in the N-terminal position. Proline peptidases PepR, PepI and PepQ act on proline containing dipeptides while PepP prefers proline containing tripeptides. The activity of these enzymes is determined by the position of the proline residue (Savijoki et al., 2006).

Many of the peptidases (e.g. PepC, PepN, PepM, PepX, PepQ) are encoded in all LAB genomes, indicating their essential role in bacterial growth, while some LAB genomes carry two peptidase homologs, most likely with identical function (Liu et al., 2010). Several paralogous genes encode other important peptidases present in all LAB genomes, such as endopeptidase PepO and dipeptidase PepV. However, aminopeptidases PepE, PepG, PepR, PepI and PepL are absent in lactococcal and streptococcal species (Liu et al., 2010).

1.1.2.4 Amino acid catabolism

The catabolism of amino acids by LAB serves a variety of physiological roles, including regulation of intracellular pH, generation of metabolic energy and resistance to stress (Fernandez and Zuniga, 2006). Furthermore, the products of amino acid catabolism by LAB contribute greatly to the characteristic flavor of various fermented products (van Kranenburg et al., 2002; Ardo, 2006; Liu et al., 2008). However, some of the produced molecules, such as biogenic amines, may adversely affect consumers' health, raising safety concerns about fermented foods (Spano et al., 2010).

The ability of LAB to degrade amino acids varies significantly among species and depends on the activity of the relevant enzymes involved in this process, like aminotransferases, decarboxylases and dehydrogenases. Methionine, leucine and phenylalanine aminotransferase activities have been detected in several lactobacilli like *Lb. paracasei*, *Lb. curvatus*, *Lb. brevis*, *Lb. plantarum* and *Lb. rhamnosus* (Williams et al., 2001). Furthermore, isolates of the genera *Lactococcus*, *Lactobacillus* and *Leuconostoc* showed aminotransferase activity towards phenylalanine and isoleucine. The highest values for both aminotransferase activities were observed for lactococcal strains (Fernández de Palencia et al., 2006). In these strains, branched-chain aminotransferase (BcAT) displays an activity towards BCAA and methionine, while aromatic aminotransferase (ArAT) is active against aromatic amino acids, leucine, and methionine. BcAT ortholog has been identified in several lactococci, streptococci and lactobacilli, but is absent in *Lb. johnsonii*, *Lb. sakei* and *Lb. reuteri*. Furthermore, putative *araT* was found in all LAB genomes except for *Lb. sakei* and *Lb. brevis* (Liu et al., 2008).

Asparagine (Asn) degradation generally starts by the action of asparaginases, which catalyze the hydrolysis of Asn to aspartate and ammonia. Studies on the available LAB genome sequences show that genes encoding putative asparaginases are present in all completed genomes and it seems that several LAB can metabolize Asn (Fernandez and Zuniga, 2006). Glutamine (Gln) is utilized by LAB for the production of glutamate and pyroglutamate through the action of glutaminase and glutaminyl cyclase, respectively (Fernandez and Zuniga, 2006). Cereal-associated *Lactobacillus sanfranciscensis* and *Lb. reuteri* (Vermeulen et al., 2007), as well as *W. cibaria* isolated from fermented pork sausage (Thongsanit et al., 2009) have been found to display glutaminase activity. Pyroglutamate has been found at high levels in long ripened cheeses (Mucchetti et al., 2000). It has been shown that pyroglutamate formation depends on the Gln cyclase activity of the thermophilic LAB used as starters (e.g *Lb. bulgaricus*, *Lb. lactis*, *Lb. helveticus*, *S. thermophilus*; Mucchetti et al., 2002).

For aspartate (Asp), LAB can employ one of the three enzymes involved in its catabolism, namely aspartate aminotransferase, aspartate decarboxylase and aspartate ammonia-lyase (Fernandez and Zuniga, 2006). Aspartate aminotransferase (AspAT) catalyzes the interconversion of aspartate and α-oxoglutarate to oxaloacetate and glutamate (Glu) by the reversible transfer of an amine group (Fernandez and Zuniga, 2006). This gene has been identified in *Lb. plantarum*, *Lb. salivarius*, *Lb. brevis*, *Lb. sakei*, *P. pentosaceus*, *S. thermophilus*, *L. lactis* subsp. *lactis*, *L. lactis* subsp. *cremoris* and *Lc. mesenteroides*, while it is absent in *Lb. acidophilus*, *Lb. johnsonii*, *Lb. gasseri*, *Lb. bulgaricus*, *Lb. reuteri*, *Lb. sakei* and *O. oeni* (Liu et al., 2008). Asp decarboxylase catalyzes the β-decarboxylation of Asp, resulting in the production of alanine (Ala) and CO_2. The initial step of this reaction includes the exchange of Asp for Ala by an aspartate:alanine antiporter and the subsequent decarboxylation of the internalized Asp (Fernandez and Zuniga, 2006). In *Tetragenococcus halophilus* genes encoding Asp decarboxylase and its related antiporter have been identified on a 25 kb plasmid (Abe et al., 2002). Finally, aspartate ammonia-lyase catalyzes the reversible conversion of Asp to fumarate acid and ammonia. The respective gene (*ansB*) has been identified in *Lb. plantarum* (Doroshchuk et al., 2006) and homologs have been found through blastn searches in the complete genomes of *Lb. paraplantarum* and *Lb. pentosus*.

The metabolism of sulfur-containing amino acids methionine and cysteine in LAB is intricative and genome-wide analysis revealed large differences in the distribution of the respective enzymes (Liu et al., 2008). *Lb. plantarum* and *S. thermophilus* strains harbor most of these genes while in other LAB species, namely *Lb. johnsonii*, *Lb. gasseri*, *Lb. brevis*, *Lb. sakei*, *P. pentosaceus*, many of these enzymes are absent. Variation in the distribution of these enzymes has been observed among strains of the same species due to absence or inactivation of genes, as in the case of *Lb. bulgaricus* for genes *metE* (homocysteine methyltransferase) and *metA* (homoserine *O*-succinyltransferase), and *L. lactis* subsp. *cremoris* for genes *cbl/cgl* (cystathionine beta lyase/cystathionine gamma lyase)*, cgs* (cystathionine gamma synthase) and *metA* (Liu et al., 2008).

The deamination of Glu by glutamate dehydrogenase (GDH) results in the production of α-oxoglutarate and ammonia (Fernandez and Zuniga, 2006). High GDH activity has been reported for facultative heterofermentative lactobacilli and leuconostocs (Fernández de Palencia et al., 2006). Furthermore, LAB presenting glutamate dehydrogenase and high aspAT activities are expected to produce acetoin or diacetyl during cheese ripening at pH value close to 5 (Le Bars and Yvon, 2008). *gdh* genes have been only found in the genomes of *Lb. plantarum*, *Lb. salivarius*, *S. thermophilus* and *E. faecalis*, while they are lacking in *Lb. acidophilus*, *Lb. johnsonii*, *Lb. gasseri*, *Lb. bulgaricus*, *Lb. brevis*, *Lb. reuteri*, *Lb. sakei*, *Lb. casei*, *P. pentosaceus*, *O. oeni*, *Lc. mesenteroides*, *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* (Fernandez and Zuniga, 2006; Liu et al., 2008). Moreover, some LAB (e.g. *L. lactis* subsp*. cremoris*, *S. thermophilus*, *Lb. brevis*) can produce γ-aminobutyrate (GABA) by the decarboxylation of Glu. The related genes *gadB* and *gadC*, encoding γ-aminobutyrate decarboxylase (GAD) and glutamate/GABA antiporter, respectively, are organized in an operon (*gad*) (Fernandez and Zuniga, 2006; Linares et al., 2016; Wu and Shah, 2017).

Serine (Ser) is employed by LAB for the production of pyruvate and NH_3 through Ser dehydratase activity (Fernandez and Zuniga, 2006). In *Lb. plantarum* serine catabolism resulted in the production of formate, succinate and acetate (Liu et al., 2003b; Skeie et al., 2008). In *P. pentosaceus* serine dehydratase activity was correlated with the presence of *dsdA* (Irmler et al., 2013), while in *Lb. sakei*, under limited glucose availability, production and activity of serine

dehydratase increased (McLeod et al., 2017). The extensive distribution of this enzyme among LAB indicates the importance of Ser catabolism for LAB physiology (Fernandez and Zuniga, 2006). In LAB, threonine (Thr) can be deaminated to 2-oxobutanoate, the precursor compound for the biosynthesis of BCAA by Thr ammonia-lyase. The respective encoding gene (*ilvA*) have been found in *L. lactis*, streptococci and *Lc. mesenteroides*. However, the major catabolic pathway of Thr in LAB is its degradation to acetaldehyde and Gly by the Thr aldolase activity Fernandez and Zuniga, 2006). Enzymatic activity has been reported for many *Lb. bulgaricus* and *S. thermophilus* strains (Ott et al., 2000). Thr aldolase activity in LAB is attributed to serine hydroxymethyltransferase (SHMT; Fernandez and Zuniga, 2006).

Histidine (His) catabolism is catalyzed by histidine decarboxylase (HDC), which results in the production of histamine in foods (Spano et al., 2010; Pessione and Cirrincione, 2016). The respective enzyme has been identified in strains of *Lb. sakei*, *Lb. curvatus*, *Lactobacillus parabuchneri*, *Lactobacillus buchneri*, *Lb. plantarum*, *Lb. brevis*, *Lb. casei*, *Lb. paracasei*, *Lb. reuteri*, *Lb. hilgardii*, *Lb. rhamnosus*, *Lactobacillus rossiae*, *Lb. helveticus*, *S. thermophilus*, *E. faecium*, *E. faecalis*, *O. oeni*, *Pediococcus parvulus*, *Lc. mesenteroides*, *W. cibaria*, *W. confusa*, *Tetragenococcus muriaticus* and *T. halophilus* (Barbieri et al., 2019). For *Lb. hilgardii, P. parvulus*, and *O. oeni*, it has been demonstrated that in the presence of histidine the expression of HDC is induced, whereas the opposite was observed upon accumulation of histamine (Landete et al., 2006). HDC encoding gene (*hdcA*) constitutes an operon together with a downstream gene (*hdcB*) of unknown function, and an upstream gene (*hdcP*), which encodes the His/histamine antiporter (Fernandez and Zuniga, 2006; Calles-Enríquez et al., 2010). The induction of this gene cluster probably confers protection to bacterial cells under acid stress conditions (De Angelis and Gobetti, 2011).

The main Arg catabolic pathway in many LAB is the arginine deiminase (ADI) pathway. The ADI pathway includes three reactions catalyzed by arginine deiminase (ADI) encoded by *arcA*, ornithine carbamoyl-transferase (OTC) encoded by *arcB*, and carbamate kinase (CK) encoded by *arcC*, and leads to the production of ornithine, ammonia and CO_2. ADI degrades Arg to citrulline and ammonia. Citrulline is the substrate of OTC and it is cleaved into ornithine and carbamoylphosphate. Finally, carbamoyl-phosphate is utilized by CK to phosphorylate ADP. The

resulting carbamate spontaneously splits into ammonia and CO_2. The ADI pathway has been described in strains belonging to the genera *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Oenococcus*, *Streptococcus*, and *Weissella* (Fernandez and Zuniga, 2006). The ability of degrading Arg has been correlated with the presence of *arc* genes (*arcA*, *arcB*, *arcC*) in strains of *Lb. hilgardii*, *Lb. brevis*, *Lb. buchneri*, *Lb. plantarum*, *O. oeni*, *P. pentosaceus*, *P. parvulus* and *Lc. mesenteroides* (Araque et al., 2009). Some LAB can also degrade Arg by decarboxylation towards the production of agmatine. Although no putative Arg decarboxylase has been identified in LAB genomes several LAB can utilize agmatine via the agmatine deiminase pathway, which includes three reactions catalyzed by agmatine deiminase (*aguA1*), putrescine carbamoyltransferase (*aguB*), and carbamate kinase (*aguC*; Fernandez and Zuniga, 2006). An agmatine deiminase operon has been identified in *L. lactis*, *Lb*. sakei, *P. pentosaceus*, *Lb. brevis*. The operon includes a gene (*aguD*) encoding an agmatine/putrescine antiporter (Lucas et al., 2007).

Furthermore, ornithine and agmatine serve as precursor compounds for the production of the biogenic amine putrescine, through the activity of ornithine decarboxylase and agmatine deiminase, respectively. The former has been identified in strains of *Lb. sakei*, *Lb. curvatus*, *Lb. buchneri*, *Lb. plantarum*, *Lb. brevis*, *Lb. paracasei*, *Lb. rhamnosus*, *Lb. rossiae*, *L. lactis*, S. *thermophilus, E. faecium*, *E. faecalis*, *Enterococcus durans*, *Enterococcus hirae*, *P. parvulus*, *O. oeni* and *T. halophilus*, while the latter in strains of *Lb. curvatus*, *Lb. plantarum*, *Lb. brevis*, *S. thermophilus*, *L. lactis*, *O. oeni*, *P. parvulus*, *P. pentosaceus*, *Lc. mesenteroides*, *E. faecalis*, *E. faecium*, *E. durans*, *E. hirae*, *Weissella halotolerans*, *Carnobacterium divergens*, *Carnobacterium maltaromaticum* and *Carnobacterium gallinarum* (Barbieri et al., 2019).

The information about the catabolic pathways for glycine (Gly), lysine (Lys), alanine (Ala) and proline (Pro) in LAB is limited. Some strains of LAB can metabolize Gly and Lys in the presence of α-ketoglutarate (Tammam et al., 2000; Williams et al., 2001). Lys decarboxylation results in the production of the biogenic amine cadaverine and the respective enzyme (i.e. lysine decarboxylase) has been found in strains of *Lb. curvatus*, *Lb. brevis*, *Lb. casei*, *Lb. paracasei*, *S. thermophilus*, *E. faecium*, *E. faecalis*, *Pediococcus* spp., *Lc. mesenteroides* and *T. halophilus* (Barbieri et al., 2019). Similarly, Ala utilization by LAB takes place either in the presence or

absence of α-ketoglutarate (Williams et al., 2001; Liu et al., 2003a). Although Ala transamination results in pyruvate formation, activity of Ala aminotransferase has not been reported in LAB (Fernandez and Zuniga, 2006). Ala racemase activity and the respective *alr* gene have been characterized in *Lb. plantarum* (Hols et al., 1997), *Lb. reuterii* (Thompson et al., 2002), and more recently in *Lb. salivarius* (Kobayashi et al., 2015). It has been demonstrated that Ala racemase is essential for cell wall biosynthesis and growth of *Lb. plantarum* (Hols et al., 1997; Palumbo et al., 2004) and *L. lactis* (Steen et al., 2005), thus, its role in Ala catabolism is probably feeble. Finally, Pro might be utilized by pyrroline-5-carboxylate reductase. Homologs of this enzyme have been found in *E. faecalis*, streptococci, *L. lactis*, and *Lb. plantarum*, but not in *Lb. acidophilus* and *Lb. johnsonii* (Fernandez and Zuniga, 2006). In some LAB (e.g. *P. pentosaceus*, *Tetragenococcus halophila*) the accumulation of proline can serve as osmoprotectant under osmotic stress (Baliarda et al., 2003). However, this is feasible only when the strain carries genes encoding the respective Pro transporters. For example, in *Lb. bulgaricus* due to the absence of genes coding for Pro transporters, like those identified in *L. lactis*, exogenous Pro was unable to alleviate stress under osmotic conditions (Chun et al., 2012).

1.1.2.5 Lipid catabolism

The lipolytic and esterolytic systems of LAB are involved in lipolysis of triglycerides and hydrophilic esters and contribute in the development of flavors in a variety of fermented foods. Lipase activity has been reported for a large number of LAB species, including *Lb. plantarum*, *Lb. acidophilus, Lb. pentosus, Lc. mesenteroides*, *E. faecium, E. faecalis, Pediococcus lolii* (Katz et al., 2002; Tanasupawat et al., 2015). Recently, a *Lb. plantarum* halotolerant lipase, active under NaCl concentrations higher than 20%, has been identified and characterized (Esteban-Torres et al., 2015). Furthermore, several studies have demonstrated that LAB species, like *O. oeni* (Sumby et al., 2009), *S. thermophilus* (Liu et al., 2001), *L. lactis* subsp. *lactis* (Chich et al, 1997; Tsakalidou and Kalantzopoulos, 1992), *Lb. casei* (Castillo et al., 1999), *Lb. plantarum* (Gobbetti et al., 1997a), and *E. faecium* (Tsakalidou et al., 1994a), possess only intracellular esterases. As a consequence, these enzymes are able to hydrolyze lipids only after the lysis of the bacterial cells and their

release in the food matrix (Chen et al., 2017). Extracellular esterases have been isolated only from *Lb. fermentum* (Gobbetti et al., 1997b).

Despite the presence of lipases and esterases, LAB, especially *Lactococcus* and *Lactobacillus* spp., are generally considered to be weakly lipolytic (Collins et al., 2003). Among thermophilic LAB, *S. thermophilus* exhibits *in vitro* significant intracellular lipolytic activity, whereas lactobacilli have weak lipolytic activity (Gobbetti et al., 1996). In addition, thermophilic lactobacilli possess a complex esterolytic system which specifies towards short chain fatty acids (El Soda et al., 1986).

Lipolysis leads to the production of various aromatic compounds. The free fatty acids (FFAs) released can be transformed to alcohols, methyl-ketones, esters and lactones (Figure 5).

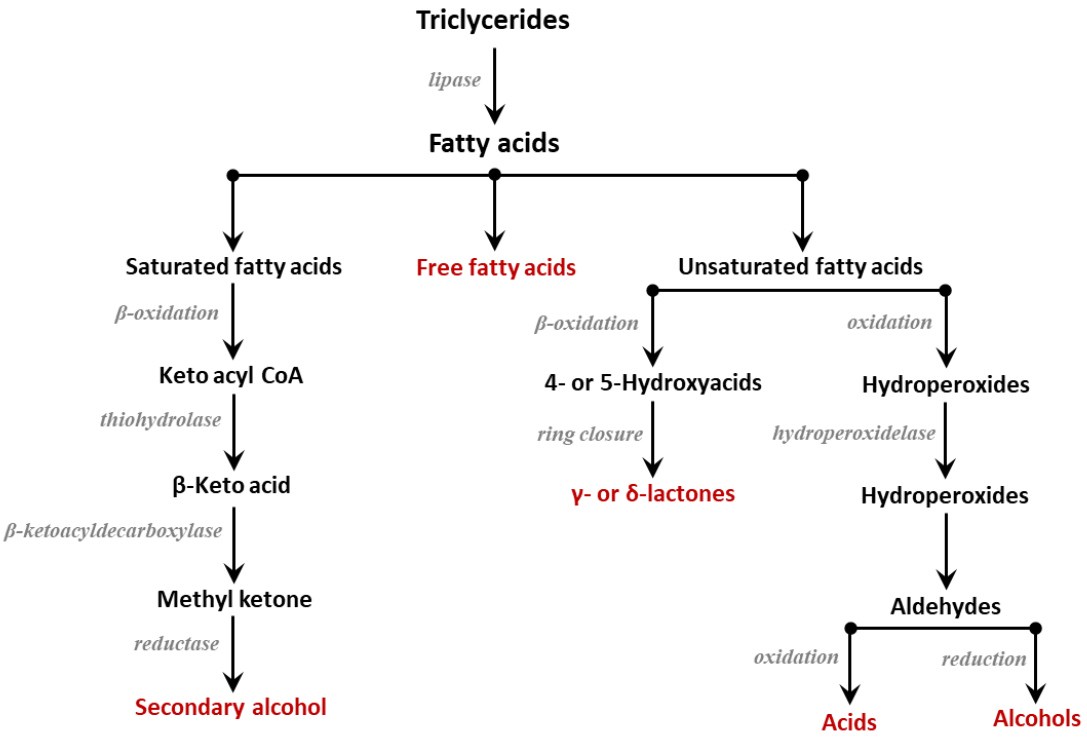


Figure 5. General degradation pathways of lipids during milk fermentation by LAB (Chen et al., 2017)

In cheese, LAB are likely considered to be responsible for the liberation of significant levels of FFA due to their presence in cheese at high numbers over an extended ripening period, despite their deficient lipolytic activity. However, excessive lipolysis is undesirable for certain cheese varieties like Gouda, Gruyere, Cheddar and Emmental, since the accumulation of volatile FFAs

may be the cause of rancid defect. In the case of hard Italian cheeses, FFA are significant contributors to the flavor. Furthermore, it has been shown that reduced-fat Cheddar cheese lacks typical flavor and contains lower concentrations of FFA, indicating the contribution of FFAs to the flavor of this cheese (Collins et al., 2003).

The release of secondary metabolites resulting from lipolysis, i.e. methyl ketones, lactones, esters and secondary alcohols, is of great importance to cheese flavor. Lactones have been reported to contribute a buttery character in cheese (Dirinck and De Winne, 1999) and have been found in several cheeses, such as Cheddar (Singh et al., 2003), Gouda (Alewijn et al., 2005), Gruyere (Rychlik and Bosset, 2001), Parmesan (Barbieri et al., 1994) and Grana Padano (Lazzi et al., 2016). Alewijn et al. (2007) studied the mechanism of lactone formation during the normal ripening of Gouda cheese concluding that was one-step, non-enzymatic reaction, where a hydroxy fatty acid, esterified in a triglyceride, undergoes trans-esterification to release the lactone directly. Furthermore, esters and thioesters are commonly found volatiles in cheese. Esters are important for development of the characteristic "fruity" flavor notes in cheeses and are formed when FFA react with alcohols (Collins et al., 2003). High concentrations of ethyl butanoate have been found in cheeses, such as Gruyere, Parmesan and Proosdij (Engels et al., 1997). In Parmesan and Grana Padano cheeses, esterases of starter and non-starter LAB seem to play a central role in the formation of ethyl butanoate and ethyl hexanoate. Purified esterases from *Lb. helveticus*, *Lb. casei*, and *L. lactis* preferentially accumulated ethyl butanoate, ethyl pentanoate, and ethyl hexanoate in a model system simulating Parmesan cheese ripening conditions. In that study it was also demonstrated that ester accumulation from LAB esterase activity was influenced greatly by a_w (Fenster et al., 2003c). S-methyl thioesters contribute a characteristic strong flavor to various smear-ripened soft cheeses (Collins et al., 2003).

The esterase encoding *estA*, which can catalyze the biosynthesis of esters of short-chain fatty acids, has been identified in many LAB genomes (e.g. *Lb. gasseri*, *Lb. plantarum*, *Lb. brevis*, *Lb. sakei*, *Lb. casei*, *O. oeni*, *Lc. mesenteroides*, *L. lactis* subsp. *lactis*, *L. lactis* subsp. *cremoris*, *S. thermophilus*). The respective gene was not present in *Lb. acidophilus*, *Lb. johnsonii*, *Lb. bulgaricus*, *Lb. reuteri*, *Lb. salivarius* and *P. pentosaceus*, indicating potential inability of the examined strains to synthesize short-chain fatty acid esters (Liu et al., 2008). *estA* has also been

found in *Lb. helveticus* (Fenster et al., 2000; Slattery et al., 2010), whereas in *Lb. casei* two additional esterase genes, namely *estB* and *estC*, have been characterized (Fenster et al., 2003b; Fenster et al., 2003a).

The contribution of lipolysis to the flavor of yogurt is limited, compared to that in long-ripened cheeses (Collins et al., 2003; Chen et al., 2017). The principal cyclic compounds in yogurt are stable γ- and δ-lactones with strong fruity flavor (Chen et al., 2017). On the other hand, the lipolytic activity of LAB affects positively the production of non-dairy fermented products. Lipids of wheat and rye flours contribute significantly to bread quality. During flour storage unsaturated fatty acids are subjected to autoxidation and subsequently they are further oxidized by cereal lipoxygenase activity during dough mixing, resulting in the production of aldehydes, the key aroma compounds in wheat and rye bread. The concentrations of these flavor compounds are significantly reduced during sourdough fermentation due to the activity of *Lb. sanfranciscensis* and *Lb. reuteri*, which convert the aldehydes to the corresponding alcohols with a much higher odor threshold (Gänzle et al., 2007). In fermented meat products the lipolytic activity has been attributed to *Lb. pentosus*, *P. lolii* (Tanasupawat et al., 2015), *Lb. curvatus* and *Lactobacillus sake* (Hammes et al., 1990).

1.1.2.6 Exopolysaccharide production

Exopolysaccharides (EPSs) are long-chain polysaccharides comprising repeating units of sugars, mainly glucose, galactose and rhamnose, in different ratios. These molecules are either secreted into the food matrix detaching permanently from the bacterial cells or remain attached on cell surface, constituting capsular polysaccharides (CPSs; Welman and Maddox, 2003). The production of CPSs has been studied for strains of *S. thermophilus*, *Lb. plantarum* and *Weissella confusa* (Low et al., 1998; Francois et al., 2004; Malang et al., 2015). Biosynthesis of EPS depends on the intermediate product of sugar degradation, namely glucose-6-phosphate, which contributes to the formation of fructose-6-phosphate employed in the glycolytic pathway, as well as the biosynthesis of sugar nucleotides, which serve as precursors for the production of EPSs molecules. The genes encoding EPS synthesis are organized adjacently, forming a gene cluster, which is located either on the chromosome, for the thermophilic strains of *Streptococcus* and

Lactobacilli or on plasmids for mesophilic LAB strains, like *L. lactis* and *Lb. casei* subsp. *casei* (Vescovo et al., 1989; van Kranenburg et al., 1997; Welman and Maddox, 2003).

S. thermophilus EPS biosynthetic gene cluster has been extensively analyzed (Bai et al., 2016; Li et al., 2018; Padmanabhan et al., 2018; Xiong et al., 2019). A recent study on the organization of EPS gene cluster among different *S. thermophilus* strains revealed enormous genetic diversity of these loci, which obviously influence the structure and the composition of the produced EPS molecules. More specifically, EPS gene cluster comprises highly conserved *epsA*, *epsB*, *epsC*, and *epsD*, which regulate the production and the elongation of EPS molecules. The number and the type of the rest adjacent *eps* genes, implicated in the formation of the repeating units *(e.g. epsE*, *epsF*, *epsG*, *epsH*, *epsI*), polymerization and export of EPS (e.g. *epsK*, *epsL*, *epsM*), vary significantly among the different gene clusters (Cui et al, 2016; Cui et al, 2017).

In *Lb. bulgaricus*, the genes for EPS biosynthesis are located on the chromosome. The respective genes have been probably acquired from *S. thermophilus* by horizontal gene transfer (HGT), due to the protocooperation of these starter microorganisms during yogurt production (Liu et al, 2009). As it has been reported by Lamothe et al. (2002), the EPS gene cluster of *Lb. bulgaricus* includes 14 genes (*epsA* to *epsN*) and the genetic organization of the loci was found to be similar to that of other *eps* clusters from LAB, such as *S. thermophilus*, *S. macedonicus* and *L. lactis* (Jolly and Stingele, 2001; Lamothe et al., 2002). Furthermore, EPS biosynthesis genes and/or the respective produced EPSs have been studied for a number of LAB species, such as *L. lactis* (Boels et al., 2003; Suzuki et al., 2013), *L. lactis* subsp. *cremoris* (Dabour and Lapointe, 2005), *Lb. rhamnosus* (Péant et al., 2005; Lebeer et al., 2009), *Lb. plantarum* (Francois et al., 2004), *Lactobacillus paraplantarum* (Zivkovic et al., 2015), *Lb. johnsonii* (Dertli et al., 2013), *O. oeni* (Dimopoulou et al., 2014), *Lb. curvatus*, *Leuconostoc lactis* (Palomba et al., 2012), *W. cibaria* and *W. confusa* (Malang et al., 2015). These gene clusters present variability in length and gene composition but also homology to proteins identified in other LAB species.

The physiological role of EPSs on the ecology of LAB is not yet completely understood. EPSs are thought to protect the bacterial cells under stress conditions related to temperature, pH, or osmotic stress (Donot et al., 2012). Furthermore, they can be involved in biofilm formation

and adhesion to surfaces (Ruas-Madiedo et al., 2002; Broadbent et al., 2003b). From the technological point of view, EPSs produced by LAB affect significantly the rheology and texture of fermented milk products, such as yogurt. It has been demonstrated that yogurt fermented with EPS-producing strains of *S. thermophilus* and *Lb. bulgaricus* showed increased mouth thickness and ropiness, high creaminess and low gel firmness (Folkenberg et al., 2006). Except for the food related technological merits, several health benefits have been attributed to EPSs produced by LAB. These include the stimulation of the immune system, antitumoral effects and decrease of blood cholesterol (Hidalgo-Cantabrana et al., 2012; Matsuzaki et al., 2014; Ryan et al., 2015; Rahbar Saadat et al., 2019). In addition, due to their structure, several EPSs present prebiotic properties which in many instances may enhance the probiotic potential of the producer strains (Zannini et al., 2016).

1.1.2.7 CRISPR-Cas systems

Clustered regularly interspaced short palindromic repeats (CRISPRs)-CRISPR associated (Cas) systems are defense mechanisms widely distributed in prokaryotes, providing acquired immunity against foreign genetic elements like viruses and plasmids (Horvath and Barrangou, 2010). CRISPR loci consist of short and highly conserved direct repeats, separated by short variable DNA sequences (spacers), flanked by diverse CRISPR-Cas genes (Horvath et al., 2009; Makarova et al., 2015). CRISPR–Cas immunity involves three distinct stages, i.e. adaptation, expression and interference (Figure 6). Adaptation comprises the incorporation of foreign DNA fragments (protospacers) from invading genetic elements into the CRISPR array as new spacers, which provide memory to the host for targeted defense against subsequent invasions by the corresponding invaders. During expression, the CRISPR array is initially transcribed as a precursor transcript (pre-crRNA), which is subsequently processed and matured to generate short CRISPR RNAs (crRNAs). During the final stage of interference, crRNAs, aided by Cas proteins, function as guides to specifically target and cleave the nucleic acids of related viruses or plasmids (Westra et al., 2014; Makarova et al., 2015). CRISPR loci provide a historical perspective of phage exposure and insight into the co-directed evolution of the phage and the host genomes, due to their role in the adaptation and persistence of a microbial host in a particular ecosystem where viruses are

present (Horvath et al., 2009). Thus, spacers within CRISPR arrays have been employed for assessing diversity among strains of the same species (Horvath et al., 2008; Delorme et al., 2017). Although the primary role of these systems is to allow microbial populations to overcome phage predation, they are also involved in the modulation of other processes, such as genetic regulation of stress responses, DNA repair and genome evolution (Sampson and Weiss, 2014; Westra et al., 2014). The genetic heterogeneity among the identified CRISPR-Cas systems is vast. Recently, Makarova et al. (2020), by combining the analysis of signature protein families and features of the architecture of *cas* loci, presented a new classification scheme for CRISPR-Cas systems, which comprises two classes, six types and 33 subtypes (Figure 6).

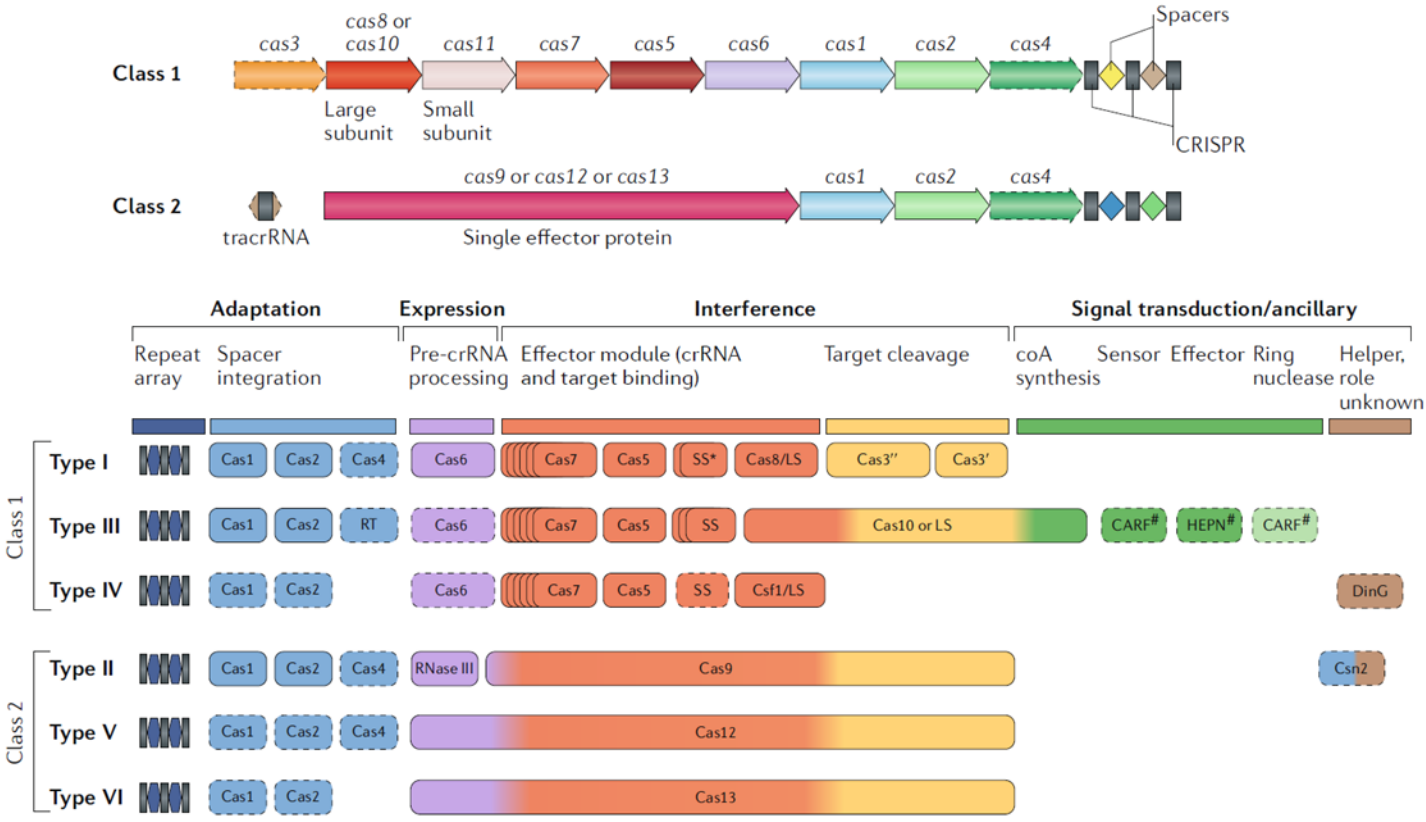


Figure 6. The two classes of CRISPR–Cas systems and their modular organization (Makarova et al., 2020)

CRISPR-Cas systems loci have been identified in several LAB, such as *Streptococcus* and *Lactobacillus*. For selected species with multiple genome sequences available, such as *E. faecalis* and *S. thermophilus*, the presence and number of CRISPR loci varied between strains (Horvath et

al., 2009). In several cases, more than one CRISPR locus could be detected within a genome. For example, in *S. thermophilus* four distinct CRISPR-Cas loci, i.e., CRISPR1, CRISPR2, CRISPR3, and CRISPR4 have been identified. The first two belong to Class 2/subtype II-A CRISPR-Cas systems, while the third and the fourth belong to Class 1/subtype III-A and Class 1/subtype I-E CRISPR-Cas systems, respectively (Horvath et al, 2008). Among them, the most prevalent in *S. thermophilus* genomes was CRISPR1, with variable size and number of spacers. *In silico* analysis of CRISPR-Cas systems in different *Lactobacillus* species revealed that the Class1/ subtype I-E and Class 2/subtype II-A were the dominant systems across the genus. In several *Lactobacillus* species (e.g. *Lb. parabuchneri*, *Lactobacillus jensenii*, *Lactobacillus ruminis*, *Lactobacillus linderi*, *Lactobacillus mucosae*, *Lb. pentosus*, *Lactobacillus animalis*, *Lb. buchneri*) the CRISPR-Cas systems are ubiquitous, while in others (e.g. *Lb. plantarum*, *Lb. reuteri*, *Lb. sakei*, *Lb. gasseri*, *Lactobacillus gallinarum*, *Lactobacillus paralimentarius*) scarce. Furthermore, some species (e.g. *Lb. acidophilus*, *Lactobacillus backii*) hold a CRISPR array in the absence of *cas* genes, whereas in other species (e.g. *Lactobacillus paracollinoides*) the opposite is observed (i.e. presence of *cas* genes and absence of CRISPR array; Crawley et al., 2018). While the CRISPR loci in the majority of studied LAB are located on the chromosome, there are cases where the respective locus is plasmid-encoded, as it has been reported for *L. lactis* (Millen et al., 2012) and *E. faecium* (Tomita and Ike, 2005).

In the case of enterococci, like *E. faecalis* and *E. faecium*, genome analysis revealed the presence of CRISPR1 and CRISPR3 loci (Class 2/subtype II-A CRISPR-Cas system) among *E. faecalis* strains to a variable degree, while an orphan CRISPR locus occurred in all *E. faecalis* strains analyzed. A CRISPR1 (Class 2/subtype II-A CRISPR-Cas system) locus was also identified in several *E. faecium* strains. In these bacteria, the absence of CRISPR-Cas systems has been associated with the emergence of antibiotic resistance in several strains of the genus. Thus, CRISPRs probably mitigate the dissemination of mobile genetic elements, which account for up to 25% of pathogenic *E. faecalis* (Palmer and Gilmore, 2010). Similarly, a negative correlation has been reported between the occurrence of prophages and the number of CRISPR spacers in *Streptococcus pyogenes*. These findings suggest the potential manipulation of CRISPRs towards

the decrease in the outspread of antimicrobial resistance genes and virulence factors among pathogens (Barrangou and Horvath, 2012).

The absence of CRISPR loci from a number of LAB species (e.g. *C. maltaromaticum*, *L. lactis* subsp. *cremoris, L. lactis* subsp. *lactis*, *Leuconostoc citreum, Leuconostoc gasicomitatum*, *Leuconostoc kimchi, Lc. mesenteroides subsp. mesenteroides, O. oeni, P. pentosaceus*) has also been reported (Barrangou and Horvath, 2012; Iskandar et al., 2017). These observations may be strain-dependent and do not preclude the existence of a CRISPR-Cas system within the genome of other strains belonging to the same species (Horvath et al., 2009). Thus, additional sequenced genomes are required in order to draw solid conclusions for the presence/absence of CRISPR-Cas systems in these species/genera.

Recently, CRISPR-Cas systems have been exploited as an efficient tool for site-specific, programmable genome editing in both single cells and whole organisms in a precise manner. The most versatile genome editing tool is the endonuclease Cas9, which belongs to Class 2/Type II CRISPR-Cas system. Cas9-based editing has now been established in *Lb. reuteri* together with ssDNA recombineering, in *Lb. plantarum* with dsDNA recombineering and plasmid-based HR, and with plasmid-based HR in *L. lactis*. Cas9 has also been used for removal of large mobile genetic elements in *S. thermophilus* and *L. lactis* (Börner et al., 2019). Furthermore, it has been demonstrated that *S. thermophilus* Cas9-based RNA-guided nucleases (RGNs) can be efficiently used for specific editing of the human genome, thus being a safe alternative to *Streptococcus pyogenes* Cas9-based RGNs (Müller et al, 2016).

1.1.3 Non-starter lactic acid bacteria (NSLAB) and interactions with starter lactic acid bacteria (SLAB)

During cheese making, milk is inoculated with SLAB, e.g. mesophilic *L. lactis* and *Leuconostoc* spp., thermophilic *S. thermophilus*, *Lb. delbrueckii* and *Lb. helveticus*, towards the rapid acidification of the cheese matrix. On the other hand, NSLAB constitute the native microbiota of both artisanal and industrially produced cheeses, deriving from raw milk, natural whey cultures and cheese making environment (Gobbetti et al., 2015). NSLAB are implicated in cheese maturation and especially in flavor development (Sgarbi et al., 2013). Their significance is

reflected upon their increasing use as adjunct cultures in the production of several industrial cheeses (Blaya et al., 2018; Peralta et al., 2018). NSLAB include obligately homofermentative species (e.g. *Lactobacillus farciminis*, *E. durans*, *E. faecalis*, *E. faecium*) facultatively heterofermentative species (e.g. *Lb. casei*, *Lb. paracasei*, *Lb. plantarum*, *Lb. pentosus*, *Lb. curvatus*, *Lb. rhamnosus*, *Pediococcus acidilactici*, *P. johnsonii, P. pentosaceus*) and obligately heterofermentative species (e.g. *Lb. fermentum*, *Lb. buchneri*, *Lactobacillus parabuchneri*, *Lb. brevis*, leuconostocs; Settanni and Moschetti, 2010).

NSLAB do not grow well in milk due to the rapid depletion of lactose within a few days of ripening by SLAB and their weak proteolytic activity (Lynch et al., 1997; Gobbetti et al., 2015). Thus, they exploit diverse energy sources, such as lactate, citrate, FFAs, peptides, free amino acids (FAAs), and ribose released from lysed cells (Williams et al., 2000; Broadbent et al., 2003a). During ripening, the enrichment of cheese matrix with nutrients, due to activity of intracellular enzymes released after SLAB, assist growth of NSLAB. In addition, some NSLAB species (e.g., *Lb. rhamnosus, Lb. paracasei*) can use energy sources beyond the expected ones, such as ribose, which may favor their survival over long maturation periods (Blaya et al., 2018). This aptitude, combined with the resistance to environmental stresses, enables NSLAB to grow slowly and eventually dominate the ecosystem of ripened cheeses (Gatti et al., 2014).

At the beginning of cheese ripening, the fate of residual lactose has been associated with the dominant microbial population and the levels of salt-in-moisture in cheese. Thus, when both levels of salt-in-moisture and NSLAB are low, residual lactose is converted mainly to L-lactate by SLAB, while at high populations of NSLAB, considerable D-lactate is formed (Blaya et al., 2018). In some instances, the accumulation of lactose, galactose and citrate in the curd, due to the inability of SLAB to metabolize them, may favor the growth of NSLAB leading to cheese defects. For example, surplus of these substrates resulted in the production of undesirable CO_2 by *Lb. casei* and atypical eye formation in Swiss-type cheese (O'Sullivan et al*.,* 2016). Furthermore, selected adjuncts cultures of NSLAB (e.g. *Lb. plantarum*, *Lb. paracasei, Lb. rhamnosus)* are used for the depletion of the residual carbohydrates in cheese, with attention to over acidification (Peralta et al., 2018). During cheese ripening, both SLAB (species of *Lactococcus* and *Leuconostoc*) and NSLAB may compete for citrate in milk. It seems that the ability of nonstarter lactobacilli,

such as *Lb. paracasei*, *Lb. plantarum* and *Lb. casei,* to metabolize residual citrate is influenced by the presence of other fermentable carbohydrates in the medium. However, the use of citrate as sole energy source has been reported for strains of *Lb. rhamnosus* and *Lb. plantarum* (Peralta et al., 2018).

In comparison to NSLAB, starter cultures, along with residual chymosin and plasmin, play a greater role in protein degradation during cheese ripening (Williams et al., 2000; Gobbetti et al., 2015). Although certain NSLAB species, such as *Lb. casei*, *Lb. paracase*i, *Lb. rhamonosus* and *Lb. buchneri*, own at least one gene coding CEP, the majority of them lack this trait (Liu et al, 2010; Sun et al., 2015). Thus, several NSLAB depend on the proteolytic activity of SLAB, the majority of which carry such proteases, for meeting their requirements in essential for growth peptides and amino acids. It seems that in the presence of SLAB the proteolytic activity of NSLAB could be stimulated. For example, genes related to protein and peptide degradation in *Lb. paracasei* were highly induced when co-cultured with *L. lactis* subsp. *cremoris* during simulated cheese ripening (Desfossés-Foucault et al., 2014). However, the same study demonstrated competition for sulfur-containing amino acids (i.e. methionine and cysteine) among the two microorganisms during ripening, since *cysK* (cystathionine β-synthase) in *L. lactis* subsp. *cremoris* and *metC* (cystathionine β-lyase) in *Lb. paracasei* were found to be upregulated in comparison to mono-cultures (Desfossés-Foucault et al., 2014). Regarding amino acid catabolism, aminotransferase activity is widespread among NSLAB and lactococci, although higher levels have been reported for SLAB (e.g. *L. lactis* subsp. *cremoris*) than for lactobacilli (e.g. *Lb. paracasei*; Kieronczyk et al., 2003; Blaya et al., 2018). Moreover, NSLAB are capable of producing GDH (*gdh*), which catalyzes the interconversion of glutamate and α-ketoglutarate, whereas this property is absent in the majority of lactococci. This enzyme activity has been reported for *Lb. plantarum, Lb. rhamnosus, Lb. fermentum and Lb. casei* (Peralta et al., 2016). Cooperation between GDH^+ non-starter lactobacilli (e.g. *Lb. casei*, *Lb. paracasei* subsp. *paracasei*) and SLAB (e.g. *L. lactis* subsp. *cremoris*) probably enhance flavor formation in cheese, since the former convert amino acids to α-keto and hydroxyl acids, while the latter employ these products for the production of carboxylic acids (Kieronczyk et al., 2003; Blaya et al., 2018).

Main sources of lipolytic enzymes in cheese include milk, rennet paste, SLAB, adjuct cultures and NSLAB (Blaya et al., 2018). In pasteurized milk cheeses, LAB contribute greater to lipolysis than endogenous enzymes, e.g. the heat unstable lipoprotein lipase of milk (Blaya et al., 2018). This is feasible upon autolysis of LAB cells, since their lipolytic enzymes are located mainly intracellularly. Although lactococci and lactobacilli are characterized by weak lipolytic systems, their activity may be sufficient for the accumulation of FFA in cheese (Collins et al., 2003). Among NSLAB, lipolytic activity has been shown for strains of *Lb. casei, Lb. plantarum* (Collins et al., 2003), *Lb. pentosus, Lc. mesenteroides*, *E. faecium and E. faecalis* (Katz et al., 2002; Tanasupawat et al., 2015), while esterolytic activity has been described in *Lb. casei* (Castillo et al., 1999), *Lb. plantarum* (Gobbetti et al., 1997a), *E. faecium* (Tsakalidou et al., 1994a) and *Lb. fermentum* (Gobbetti et al., 1997b). In *Lb. casei*, *estB* and *estC* were active during simulated cheese ripening conditions, contributing significantly to the accumulation of ethyl esters (Fenster et al., 2003c). In contrast, the *estA* of *Lb. paracasei* in mixed culture with *L. lactis* subsp *cremoris* was downregulated during cheese manufacture and ripening, indicating inability of the strain to produce esters in cheese (Desfossés-Foucault et al., 2014).

During dairy products manufacture and ripening processes, both SLAB and NSLAB are exposed to unfavorable growth conditions, related mostly to temperature, acidity, osmotic pressure and nutrient availability. To overcome these hurdles and persist to this hostile environment, LAB have developed adaptive stress response mechanisms (Franz and Holzapfel, 2011). The endurance of SLAB and NSLAB under these adverse conditions can determine their prevalence in the final product. Molecular chaperones (e.g. *dnaK*, *dnaJ*, *grpE*, *groEL*, *groES*) along with protease chaperones (e.g. *htrA*, *clpB*, *clpP*, *clpX*, *fstH*) are induced under stress conditions associated with high temperatures (e.g. pasteurization), while cold shock proteins are produced at low temperatures (e.g. ripening, storage) as response to cold stress (Franz and Holzapfel, 2011; Papadimitriou et al., 2016). Among NSLAB, heat shock response has been studied for *Lb. plantarum* (De Angelis et al., 2004; Russo et al., 2012), *Lb. casei* (Adu et al., 2018), *Lb. paracasei* (Desmond et al., 2004), *Lb. gasseri* (Suokko et al., 2008) and *Lb. rhamnosus* (Prasad et al., 2003), while cold-shock response has been investigated in *Lb. plantarum* (Derzelle et al., 2003) and *Lb. casei* (Sauvageot et al., 2006).

LAB possess several mechanisms for coping with acid stress. These include the induction of *clp* chaperones, the modification of pyruvate metabolism in expense of lactate, and the synthesis of compounds, such as basic metabolites, energy-rich intermediates and EPS (e.g. *Lb. plantarum*, *Lb. reuteri*, *Lb rhamnosus*), the metabolism of citrate (e.g. *Enterococcus*, *Lb. plantarum*, *Lb. rhamnosus*) and the activity of F_1F_o-ATPase proton pump (Papadimitriou et al., 2016). The maximum activity of the latter has been observed at low pH values (5.5) for NSLAB (e.g. *Lb. casei*) and higher ones (7.5) for SLAB (e.g. *L. lactis* subsp. *cremoris*, *S. thermophilus*; Nannen and Hutkins, 1991). The activation of mechanisms involved in the production of ammonia, like the arginine deiminase pathway, the urease system (e.g *Lb. reuteri*, *Lb. fermentum*) and the production of GDH (e.g. *Lb. plantarum, Lb. rhamnosus, Lb. fermentum, Lb. casei*), can also confer protection to bacterial cells against acid stress (Franz and Holzapfel, 2011; Papadimitriou et al., 2016; Peralta et al., 2016). Furthermore, amino acid decarboxylation reactions can alleviate acidic stress, but results in the accumulation of biogenic amines in cheese (Zuljan et al., 2016). Since this trait is undesirable in industrially employed SLAB and thus must be absent, these metabolites are most probably produced by non-starter lactobacilli and enterococci (Zuljan et al., 2016).

General stress proteins (*groEL*, *groES*, *dnaK*) and proteases (*ftsH*, *htrA*) can be also induced in LAB as response to osmotic stress (Franz and Holzapfel, 2011; Papadimitriou et al., 2016). In *Lb. rhamnosus*, GroEL and DnaK heat shock proteins were upregulated due to osmotic shock treatment (Prasad et al., 2003). However, the main mechanism of osmotic stress response in LAB includes the uptake of osmoprotectants (e.g. glycine-betaine), facilitated by the ATP-dependent *qacT* system (e.g. *Lb. plantarum*; Glaasker et al., 1998) and the *opuA* or *busA* ABC transporters (*Lb. paracasei*; Desfossés-Foucault et al., 2014). Furthermore, it has been reported that proline can be employed by LAB to diminish the strength of osmotic stress (e.g. *Lb. plantarum*; Glaasker et al., 1998, *P. pentosaceus*; Baliarda et al., 2003).

Starvation stress is caused due to the limitation of the available nutrients during LAB growth. Ganesan et al. (2007) has reported that during carbohydrate starvation, lactococcal strains (i.e. *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris*) were able to survive by repressing cell division and sugar metabolism, and inducing amino acid metabolism. In *L. lactis* subsp.

cremoris during co-cultivation with *Lb. paracasei*, lactate dehydrogenase (*ldh*) and galactokinase (*galK*) were positively influenced. That was not the case for *Lb. paracasei*, indicating that the latter is less affected by carbon starvation in comparison to the former (Desfossés-Foucault et al., 2014). In *Lb. casei*, lactose starvation induced the citrate metabolic pathway and the catabolism of various carbon sources (e.g. inositol, glycerol, peptides, FFAs), aiming at cellular energy production (Al-Naseri et al., 2013). The presence of such adaptive physiological and metabolic mechanisms in NSLAB allow them not only to survive but dominate the cheese environment during ripening.

1.1.4 Applications of lactic acid bacteria

Production and consumption of fermented foods is a common practice implemented throughout the world for millennia. Even nowadays, fermentation is used as a method of production and preservation of foods in the households of several countries, where the availability of food commodities is limited. The long history of safe use of fermented foods along with the technological advancements in food processing, enabled the large-scale production of fermented products with well-designed control over the microorganisms used in the fermentation (Naidu et al., 1999). As shown in Table 2, LAB are widely employed in the industrial production of various fermented products as starters or adjunct cultures, due to their metabolic properties, which contribute to the organoleptic characteristics, the nutritional value and the microbial safety of the final products (Leroy and De Vuyst, 2004; Settanni and Moschetti, 2010; Kongo, 2013). Species widely used in industrial food fermentations belong to the genera *Lactococcus* (dairy), *Lactobacillus* (dairy, meat, vegetables, cereals), *Leuconostoc* (vegetables, dairy), *Pediococcus* (vegetables, meat), *Oenococcus* (wine), and *Streptococcus* (dairy) (Bourdichon et al., 2012). Furthermore, the technological properties of less prominent food production related species, such as *C. maltaromaticum* (Afzal et al., 2010), *T. halophilus* (Kim et al., 2019) and *Weissella cibaria* (Lynch et al., 2014), are increasingly studied with promising results for their ensuing use as starter or adjunct cultures in food industry.

Table 2. LAB species used as starter and adjunct cultures in dairy products (Kongo, 2013)

Species	Main uses
Lactococcus	
L. lactis subsp. *lactis*	Mesophilic starter for many cheese types
L. lactis subsp. *lactis biovar diacetylactis*	Gouda, Edam, sour cream, lactic butter
L. lactis subsp. *cremoris*	Mesophilic starter for many cheese types
Streptococcus	Thermophilic starter for yogurt and many cheese types,
S. thermophilus	hard and semi hard high-cooked cheeses
Lactobacillus	
Lb. acidophilus	Probiotic adjunct culture for cheese and yogurt
Lb. bulgaricus	Thermophilic starter for yogurt, many cheese types, hard and semi hard high-cook cheeses
Lb. lactis	Fermented milks and high-cook cheese
Lb. helveticus	Thermophilic starter for fermented milks and many cheese types, hard and semi hard high-cook cheeses
Lb. casei	Cheese ripening adjunct culture
Lb. plantarum	Cheese ripening adjunct culture
Lb. rhamnosus	Cheese ripening adjunct culture
Leuconostoc	Mesophilic culture for Edam, Gouda, fresh cheese, lactic butter, sour
Lc. mesenteroides subsp. *cremoris*	cream

During the last decades, LAB are increasingly used as probiotics, i.e. as "live microorganisms that when administered in adequate amounts confer a health benefit on the host" (FAO/WHO, 2001) in the production of fermented foods and food supplements. The socioeconomic significance of probiotics is attested by a recently published market report by Allied Market Research. By 2022, the global probiotics market is expected to gain $57.4 billion, registering a compound annual growth rate (CAGR) of 7.7% during the period 2016-2022 (Oyeniran et al., 2020). Several LAB species meet the requirements defined for probiotic microorganisms, as presented in Table 3, with adherence to the host cell surface, which is considered to be a prerequisite for colonization, and health-promoting effects, being among the most significant (Otero et al., 2004; Oyeniran et al., 2020). The genera *Lactococcus* (*L. lactis*), *Lactobacillus* (*Lb. bulgaricus*, *Lb. acidophilus*, *Lb. casei*, *Lb. rhamnosus*, *Lb. reu*teri, *Lb. plantarum*), *Streptococcus* (*S. thermophilus*) and *Enterococcus* (*E. faecium*, *E. faecalis*) are the most commonly studied probiotic microorganisms (Guarner et al., 2007; Yadav et al., 2007; Franz et al., 2011; Lee et al., 2015; Mantegazza et al., 2018).

Health-beneficial effects ascribed to probiotic LAB (Figure 7) include the stabilization of gut microbiota and the competitive exclusion of enteric pathogens, the immunomodulation, the

reduction of serum cholesterol by assimilation mechanisms, the control of cardiovascular function through antithrombotic and anti-hypotensive action, the decreased risk of colon cancer by detoxification of carcinogens, the suppression of tumors by modulation of cell mediated immunity and the control of gut-brain axis (Naidu et al., 1999; Divya et al., 2012; Pessione and Cirrincione, 2016). The proteolytic activity of LAB on different food proteins results in the production of several health-beneficial peptides. Alpha and beta-caseins, albumin and globulin from milk and dairy products, rubisco from spinach, beta-conglycinin from soy and gluten from cereals constitute good matrices for the production of these important bioactive compounds (Pessione and Cirrincione, 2016).

Table 3. Criteria for the selection of probiotics in commercial applications (Oyeniran et al., 2020)

Criteria	**Property**
Safety	Origin Pathogenicity and infectivity Virulence factors-toxicity, metabolic activity and intrinsic properties, i.e. antibiotic resistance
Technological	Genetically stable strains Desired viability during processing and storage Good sensory properties Phage resistance Large-scale production
Functional	Tolerance to gastric acid and juices Bile tolerance Adhesion to mucosal surface Validated and documented health effects
Health-promoting	Immunomodulation Antagonistic activity towards gastrointestinal pathogens, i.e., *Helicobacter pylori*, *Candida albicans* Cholesterol metabolism Lactose metabolism Antimutagenic and anticarcinogenic properties

The probiotic efficacy of LAB is also employed in animal husbandry. The administration of probiotics and direct-fed microbials (DFMs) to animals has gained ground over the systematic and inconsiderate use of antibiotics and presents a variety of beneficial effects. These include treatment of digestive disorders and reduction of gut pathogens, stabilization of the gastrointestinal pH, enhanced animal and growth performance due to efficient absorption of nutrients, increased feed conversion efficiency and diet digestibility, improved milk yield and

composition, stimulation of the immune system, treatment of mastitis, methane mitigation and reduction of ammonia's concertation in the excreta (Zoumpopoulou et al., 2018a).

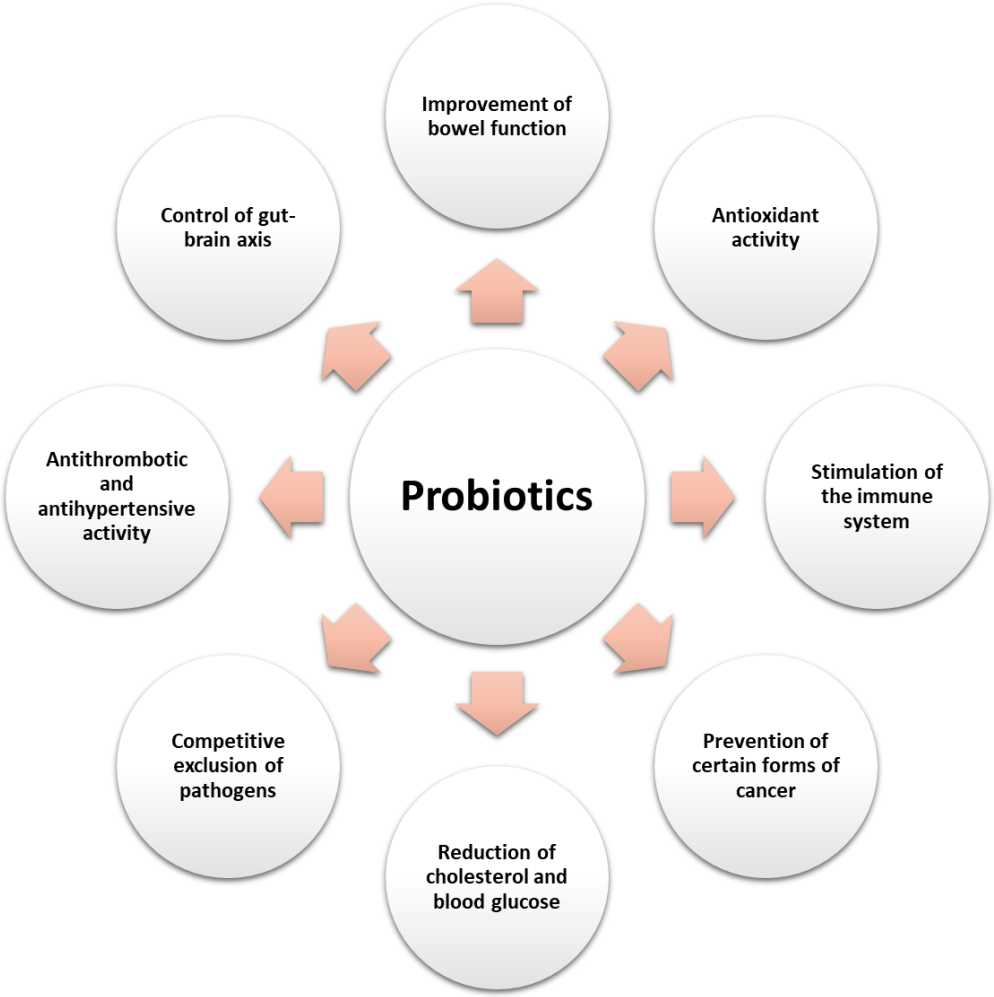


Figure 7. Health benefits of probiotics (Divya et al., 2012; Pessione and Cirrincione, 2016)

Furthermore, LAB produce a plethora of metabolites during food fermentation. The biopreservation of fermented foods is attributed to the antimicrobial activity of LAB due to the production of organic acids (lactic and acetic acid), hydrogen peroxide, ethanol, CO_2, and bacteriocins (Šušković et al., 2010; Yang et al., 2012). Bacteriocin production is a rather desirable probiotic trait, since bacteriocin-producing probiotics could be used as an alternative to antibiotics *in vivo* (Papadimitriou et al., 2014). Furthermore, bacteriocins produced by LAB can be applied in the food industry as a natural barrier against foodborne pathogens and spoilage

microorganisms, such as *Listeria monocytogenes* and *Staphylococcus aureus* (O'Sullivan et al., 2002). They are safe for human consumption since they can be easily degraded by the proteases of the mammalian gastrointestinal tract (GIT). The bacteriocin molecules are small, ribosomally synthesized peptides or proteins with antimicrobial activity against closely related Gram-positive bacteria, while producer bacteria are immune to their own bacteriocin(s). Through the action of bacteriocins the membrane of the target bacterial cells becomes permeable resulting to the leakage of ions, ATP, and other vital molecules (De Vuyst and Leroy, 2007; Zacharof and Lovitt, 2012). Bacteriocins from LAB are categorized in three distinct classes. Class I bacteriocins or lantibiotics include small, heat-stable, lanthionine-containing, mono- and di-peptide bacteriocins, like nisin. The mature bacteriocins are generated from the respective inactive precursor peptides after extensive post-translational modification. Class II bacteriocins or non-lantibiotics include small, heat-stable, non-lanthionine-containing bacteriocins, and are further subdivided to pediocin-like bacteriocins (class IIa), two-peptide bacteriocins, such as Lactacin F and lactococcin G (class IIb), and circular bacteriocins (class IIc). Finally, class III bacteriocins include large, heat-labile, lytic proteins, such as helveticin I and enterolysin, produced by *Lb. helveticus* and *E. faecium*, respectively (De Vuyst and Leroy, 2007; Zacharof and Lovitt, 2012; Alvarez-Sieiro et al., 2016).

Furthermore, important developments have been achieved in the design of recombinant LAB for mucosal delivery of therapeutic and prophylactic compounds, such as antigens, allergens, immune modulators, antimicrobial peptides, single-chain antibodies and a few enzymes (Wells and Mercenier, 2008). More specifically, in the past two decades *L. lactis* and *Lb. plantarum* have been employed as vectors for the delivery of beneficial molecules via the intranasal, oral, or genital mucosal surfaces. Other lactobacilli have been studied as live recombinant vectors for the development of new safe mucosal vaccines, including *Lb. casei*, *Lactobacillus jensenii*, *Lb. acidophilus* and *Lb. rhamnosus* (LeBlanc et al., 2013; Wang et al., 2016). The results of several studies demonstrate that this approach seems to be a realistic therapeutic option for the treatment of a wide variety of human diseases (Wang et al., 2016).

Finally, LAB, and especially species of the genus *Lactobacillus*, are also applicable in agricultural systems. More specifically, they are used in the production of composts for the

improvement of soils, in the control of bacterial and fungal phytopathogens of crop plants, for the amelioration of abiotic stress in plants and as plant growth stimulants (Lamont et al., 2017).

1.1.5 The genus *Streptococcus*

1.1.5.1 General information

The genus *Streptococcus* consists of gram-positive, spherical or ovoid, homofermentative, facultatively anaerobic, non-spore forming, mostly non-motile, catalase-negative bacteria, with a relatively low G+C content (ranging from 34 to 46%). Streptococcal cells are typically arranged in chains or pairs (Tagg et al., 2019).

The genus includes a large number of pathogens and commensals, several of which occur as physiological flora in the oral cavity and intestine of humans and animals, and they often inhabit skin, throat, and the upper respiratory tract. Several pathogenic species present increased adaptability to new hosts and resistance to antibiotics and immune responses. Additionally, these commensal streptococci occur as opportunistic pathogens causing infections due to weak immunological response of the host. The severe impact of these bacteria on human and animal health is reflected on the proliferation of the morbidity and mortality rates worldwide and the substantial economic losses for the food and agriculture sectors (Cole et al., 2008; Krzyściak et al., 2013; Gao et al., 2014; Richards et al., 2014).

Pathogenic streptococci include species commonly causing infection in humans (e.g. *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus pneumonia*), commensal species colonizing mucosal niches and the surface of teeth (e.g. *Streptococcus mitis*, *S. mutans*, *Streptococcus salivarius*, *Streptococcus anginosus*) and zoonotic species, which may cause opportunistic infection in humans (e.g. *Streptococcus dysgalactiae* subsp. *equisimilisa* and *Streptococcus equi* subsp. *zooepidemicus* from horses; members of the *Streptococcus bovis* group from cattle; *Streptococcus canis* from dogs; *Streptococcus porcinus* and *Streptococcus suis* from pigs; *Streptococcus iniae* from fish; *Streptococcus gallinaceus* from chickens). Within the genus, only *S. thermophilus* is nonpathogenic and is used extensively in the dairy industry as starter culture (Cole et al., 2008; Krzyściak et al., 2013; Gao et al., 2014; Richards et al., 2014).

The genus *Streptococcus* has been revisited taxonomically, on the basis of advancements in chemotaxonomic methods, DNA hybridization and 16S rRNA gene sequencing (Tagg et al., 2019). The genus is subdivided into six major groups (Figure 8), namely:

- the Pyogenic group, which includes basically β-hemolytic pathogenic species for humans and animals (e.g. *S. pyogenes*, *S. agalactiae*, *Streptococcus urinalis*, *S. dysgalactiae*, *S. equi*, *S. iniae*, *S. porcinus*, *S. canis*)
- the Anginosus group, which consists of commensal and occasional opportunistic pathogens found in the oral cavity and the gastrointestinal and genital tract of humans (e.g. *S. anginosus*, *Streptococcus constellatus*, *Streptococcus intemedius*)
- the Mitis group, which includes the pathogen *S. pneumoniae* and various oral commensals (e.g. *S. mitis*, *Streptococcus oralis*, *Streptococcus infantis*, *Streptococcus australis*, *Streptococcus sanguinis*, *Streptococcus cristatus*, *Streptococcus gordonii*)
- the Salivarius group, which comprises dairy streptococci and commensals of the human oral cavity (e.g. *S. thermophilus*, *Streptococcus vestibularis*, *S. salivarius*)
- the Bovis/equinus group, which contains species primarily found in the intestinal tract of several animals and in fermented dairy products (e.g. *Streptococcus infantarius*, *Streptococcus equinus*, *Streptococcus macedonicus*, *Streptococcus bovis*, *Streptococcus gallolyticus*)
- the Mutans group, which comprises genetically heterogeneous species that are nevertheless phenotypically similar (e.g. *S. mutans*, *Streptococcus ratti*, *Streptococcus criceti*, *Streptococcus downei*)

Finally, one additional group, designated as unknown, includes ungrouped species, characterized by unclear phylogenetic relationships (e.g. *S. suis*; Gao et al., 2014; Tagg et al., 2019).

Furthermore, population structure analysis revealed two distinct lineages, which include Pyogenic, Bovis, Mutans and Salivarius groups, as well as Mitis, Anginosus and Unknown groups, respectively, as shown in Figure 8 (Gao et al., 2014). Richards et al. (2014) reported that the early evolution of *Streptococcus* was characterized by greater gene gain than loss, whereas the

evolution of the major groups was characterized in general by more gene loss than gain. Furthermore, a large proportion of *Streptococcus* pan-genome has been shaped by HGT events. However, several biochemical characteristics have been retained since group formation (e.g. proteolysis in Mitis group, urea metabolism in Salivarius group, carbohydrate metabolism in Pyogenic group transcription regulation in Bovis group; Richards et al., 2014). These findings indicate not only genomic cohesiveness through time but also the importance of certain metabolic properties for each group, related to the ecological niche they occupy (Richards et al., 2014).

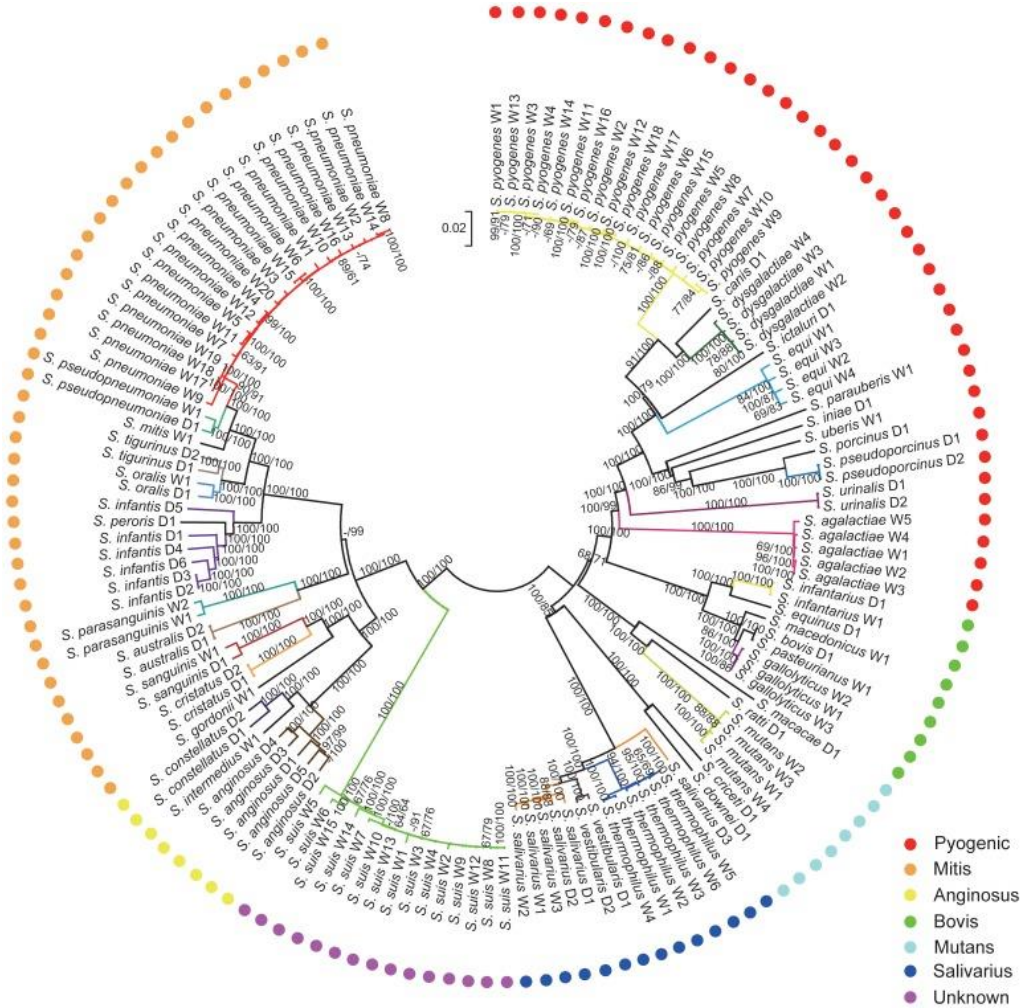


Figure 8. Phylogenomic tree of the genus *Streptococcus* (Gao et al., 2014)

1.1.5.2 *Streptococcus thermophilus*

S. thermophilus belongs to the Salivarius group, which includes also the species *S. salivarius* and *S. vestibularis*, two commensal bacteria that may occasionally cause opportunistic infections in humans. The group is a cluster shaped upon the basis of sequence analysis using the 16S (Bentley et al., 1991; Kawamura et al., 1995), *sodA* (Poyart et al., 1998), *rnpB* (Täpp et al., 2003) and, more recently, 136 (Richards et al., 2014) and 278 (Gao et al., 2014) core streptococcal genes.

S. thermophilus has been isolated mainly from yogurt (Bolotin et al., 2004; Makarova et al., 2006; Kang et al., 2012; Bai et al., 2016; Alexandraki et al., 2017a; Evivie et al., 2017; Li et al., 2018) but has been also identified in raw milk (Delorme et al., 2011; Sun et al., 2011b; Labrie et al., 2015; Renye et al., 2017) and traditional cheeses (Shi et al., 2015). Apart from dairy products, strains have been isolated from fish intestine (Linares et al., 2016; Linares et al., 2017), commercial dietary supplements and plants (Michaylova et al., 2007). *S. salivarius*, which can be isolated from human saliva and the tongue dorsum, is one of the first bacteria to colonize the mucosa of neonates in the first few days after birth. Is also one of the primary inhabitants of the human gut microbiota (Delorme et al., 2015). Finally, *S. vestibularis*, is a mutualistic bacterium present on the vestibulum of human oral mucosa (Whiley and Hardie, 1988).

S. thermophilus is the only species within the *Streptococcus* genus, which has been granted the generally recognized as safe (GRAS) status and the qualified presumption of safety (QPS) status according to the Food and Drug Administration (FDA, 2007) and the European Food Safety Authority (EFSA, 2007), respectively. The species is employed as a starter culture in the production of yogurt (Chen et al., 2017), fermented milks (e.g. kefir; Beshkova et al, 2002), and several cheese varieties (e.g. Mozzarella, Gorgonzola, Gruyere, Feta; Parente et al., 2017; Rantsiou et al., 2008). The economic value of *S. thermophilus* as a starter and the demand of the dairy industry for novel starter strains with improved functionality, motivated the genomic research of the species. Several studies yielded significant information about *S. thermophilus* technological traits and its adaptation to the milk niche.

S. thermophilus has adapted to the rich in nutrients milk habitat through extensive reductive evolution, probably during species domestication toward a starter culture, a

hypothesis, which is supported by the numerous pseudogenes identified in the complete genomes of several *S. thermophilus* strains analyzed so far (Bolotin et al., 2004; Hols et al., 2005; Goh et al., 2011; Papadimitriou et al., 2015). The loss of dispensable genes related to carbohydrate catabolism and pathogenicity has led to genome size reduction and simplification of species metabolism (Bolotin et al., 2004; Hols et al., 2005; Goh et al., 2011). The acquisition of genes related to EPS production, stress response, phage immunity and protocooperation due to HGT events, has further shaped its adaptation, metabolic and technological traits (Liu et al., 2009; Eng et al., 2011). The *gal-lac* gene clusters (*galRKTEM-lacSZ*), implicated in the fermentation of lactose and galactose, are highly conserved (Vaughan et al., 2001). Although, the majority of *S. thermophilus* strains are unable to metabolize galactose, it has been demonstrated that even Gal^- strains can slowly consume an amount of the produced galactose, while in Gal^+ strains the consumption is faster and to completion (de Vin et al., 2005). Furthermore, no GalK activity was detected in Gal^+ strains, indicating the existence of an alternative pathway for galactose catabolism by these strains (de Vin et al., 2005).

EPS biosynthetic loci in *S. thermophilus* generally comprise *epsA* and *epsB*, which are involved in regulation of EPS production, as well as *epsC*, and *epsD,* which are associated with the chain-length determination of EPS (Cui et al., 2016; Cui et al., 2017). *epsE* encodes enzymes that initiate the assembly of the EPS repeating units (Broadbent et al., 2003b; Wu et al., 2014). The genes downstream *epsE* encode proteins related to transport of nucleotide sugar moieties (i.e. *epsF*, *epsG*, *epsH*, *epsI*), as well as polymerization and translocation of the EPS (i.e. *epsK*, *epsL*, and *epsM*; Goh et al, 2011; Wu et al., 2014; Cui et al., 2016; Cui et al, 2017; Evivie et al., 2017).

Proteolysis is initiated by the cell-wall associated proteinase PrtS, which is responsible for the degradation of caseins (Hols et al., 2005; Goh et al., 2011; Tian et al., 2018). However, the respective gene is not present in all *S. thermophilus* strains (Goh et al., 2011). From a technological point of view, the presence of *prtS* is vital for the rapid growth of *S. thermophilus* in milk as a monoculture, and, thus, milk rapid acidification (Courtin et al., 2002). However, it seems that the rate of acidification depends on the induction of several genes and the sole presence of *prtS* is inadequate to ensure this property (Galia et al., 2016). In addition, it has been reported that only *prtS* deficient *S. thermophilus* strains can perform protocooperation with *Lb.*

bulgaricus (Settachaimongkon et al., 2014). The majority of peptide and amino acid transporters identified in *S. thermophilus* belong to the ABC superfamily and include, among others, oligopeptide Opp ABC transporters, branched-chain amino acid ABC transporters, amino acid ABC transporters and spermidine/putrescine ABC transporters. Furthermore, di-tripeptide transporters and amino acid symporters have been also found (Hols et al., 2005; Goh et al., 2011). The presence of a glutamate/GABA (*gadC*) and histidine/histamine (*hdcP*) antiporters have also been reported (Linares et al., 2016; Calles-Enríquez et al., 2010). Moreover, highly conserved cytoplasmic peptidases found in *S. thermophilus* include *pepA, pepC, pepF, pepM, pepN, pepO, pepP, pepQ, pepS, pepT, pepV*, and *pepX* (Liu et al., 2010).

Biosynthesis of proline, branched-chain amino acids, glutamine and aspartate has been experimentally studied in *S. thermophilus* (Limauro et al., 1996; Garault et al., 2000; Monnet et al., 2005; Arioli et al., 2007). Furthermore, amino acid omission studies demonstrated that the number and the type of essential amino acids required for growth is strain-dependent (Garault et al., 2000; Letort and Juillard, 2001; Pastink et al. 2009). Thus, some *S. thermophilus* strains exhibited no amino acid requirement, while others appeared auxotrophic for certain amino acids (Glu, Cys, His, and Met; Letort and Juillard, 2001; Pastink et al. 2009). The *in silico* analysis of amino acid biosynthetic pathways in *S. thermophilus* revealed the presence of the majority of genes involved in these, with the exception of His biosynthesis related genes, which were missing in a number of strains (Hols et al., 2005; Goh et al., 2011).

S. thermophilus also owns immunity mechanisms, such as CRISPR-Cas and R-M systems, which confer resistance against parasitic DNA (e.g. phages, plasmids; Horvath and Barrangou, 2010; Dupuis et al., 2013). Studies have shown that strains of the species carry up to four distinct CRISPR-Cas loci (Horvath and Barrangou, 2010), whereas the spacers of these CRISPR arrays have been utilized for the evaluation of diversity within the species (Horvath et al., 2008; Delorme et al., 2017).

The probiotic efficacy of *S. thermophilus* has been primarily correlated with the health benefits attributed to yogurt consumption (Freitas, 2017). Although the species is sensitive to gastric acids, human studies have demonstrated that several *S. thermophilus* strains not only can survive passage through the GIT but also adhere to intestinal epithelial cells (Uriot et al., 2017).

Health benefits associated with *S. thermophilus* include alleviation of lactose intolerance due to β-galactosidase production (Drouault et al., 2002), reduction of intestinal inflammation and improvement of mucosal barrier function (Bailey et al., 2017; Chen and Zhang, 2019), stimulation of the gut immune system due to EPS production (Mizuno et al., 2020) and anti-carcinogenic activity (Tarrah et al., 2018).

1.1.6 The genus *Lactobacillus*

1.1.6.1 General information

The genus *Lactobacillus* is a diverse group of anaerobic or aerotolerant, rod-shaped, gram-positive, non-spore forming, mostly non-motile, catalase negative bacteria. Based on their main sugar fermentation pathways (glycolysis and pentose phosphate pathways) the *Lactobacillus* species are classified as obligately homofermentative or facultatively/obligately heterofermentative (Salvetti et al., 2018). Lactobacilli are commonly found in various nutrient-rich environments, like food matrices, microbial-heavy host habitats, such as human and animal GIT, as well as in natural ecological niches, such as plants and soil (Sun et al., 2015). Their ability to ferment raw milk, meat and plant material contributes significantly to the production of artisanal and industrial fermented foods (Sun et al., 2015). Moreover, several strains are well known for their probiotic properties exhibiting health benefits beyond the basic nutritional value (Lebeer et al., 2008). Commercially relevant species include *Lb. acidophilus*, *Lb. casei*, *Lb. bulgaricus*, *Lb. fermentum*, *Lb. paracasei*, *Lb. plantarum*, *Lb. reuteri*, *Lb. rhamnosus*, and *Lb. salivarius* (Morovic et al., 2016). Based on food regulatory framework, 84 *Lactobacillus* species are certified for safe, technological and beneficial use (European Food and Feed Cultures Association, EFFCA; Bourdichon et al., 2012), 36 species have the QPS status (EFSA; Ricci et al., 2017), and 12 species have the GRAS status (FDA; Salvetti and O'Toole, 2017). The economic significance of lactobacilli used as probiotics and DFMs is enormous and by 2022 are expected to reach a value of 64 and $1.4 billion, respectively (Salvetti et al., 2018).

The genus *Lactobacillus* is currently comprising more than 250 species (http://www.bacterio.net, accessed on July 2020). The phylogenetic tree of *Lactobacillus* genus is complex, showing that lactobacilli branch in several clades (highlighted in different colors),

interspersed by other genera, such as *Pediococcus* (*Lactobacillaceae* family) and *Fructobacillus*, *Leuconostoc*, *Oenococcus* and *Weissella* (*Leuconostocaceae* family), as shown in Figure 9. The analysis revealed the relation among obligately heterofermentative lactobacilli and genera of *Leuconostocaceae* family, which display common metabolism, and their separation from the homofermentative and facultatively heterofermentative *Lactobacillus* species.

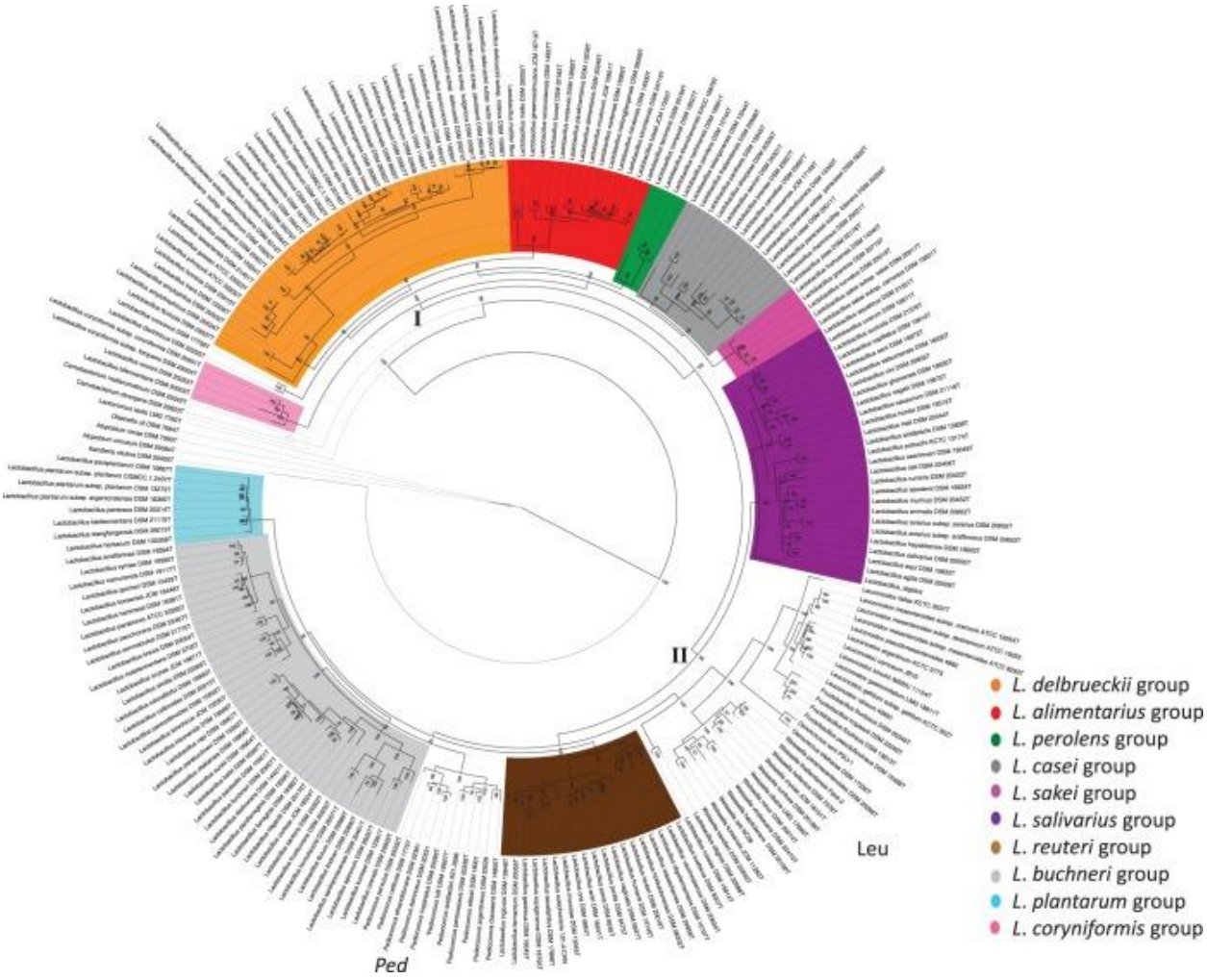


Figure 9. Phylogenetic tree for selected strains of the genus *Lactobacillus* and related genera based on the amino acid sequences of 29 ribosomal proteins (Salvetti et al., 2018)

Furthermore, 10 *Lactobacillus* groups have been defined, namely, the *Lb. delbrueckii* group (e.g. *Lb. helveticus*, *Lb. acidophilus*, *Lb. gasseri* along with the peripheral species *Lactobacillus amylophilus*, *Lactobacillus amylotrophicus*, *Lactobacillus floricola*), the *Lb. casei*

group (e.g. *Lb. paracasei*, *Lb. rhamnosus*), the *Lb. salivarius* group (e.g. *Lb. acidipiscis*, Lb*. ruminis*, *Lb. agilis*, *Lb. equi*), the *Lb. reuteri* group (e.g. *Lb. fermentum*), the *Lb. buchneri* group (e.g. *Lb. zymae*, *Lb. brevis*, *Lb. sunkii*, *Lb. kefiri*, *Lb. hilgardii*, *Lb. fructivorans*, Lb. *collinoides*), the *Lb. plantarum* group (e.g. *Lb. paraplantarum*, *Lb. pentosus*), the *Lb. coryniformis* group (e.g. *Lb. rennini*), the *Lb. alimentarius* (e.g. *Lb. paralimentarius*, *Lb. crustotum*, *Lb. farciminis*) group, *Lb. perolens* (e.g. *Lb. composti*) group and the *Lb. sakei* group (e.g. *Lb. graminis*, *Lb. curvatus*). Only five *Lactobacillus* species were not clustered within any *Lactobacillus* group, namely *Lb. rossiae*, *Lactobacillus siliginis*, *Lactobacillus concavus, Lactobacillus dextrinicus*, and *Lactobacillus selangorensis* (Salvetti et al., 2018). Recently, Zheng et al. (2020) proposed a reclassification scheme for *Lactobacillus* genus into 25 genera (*Lactobacillus*, *Paralactobacillus* and 23 novel genera), based on the evaluation of core genome phylogeny, pairwise average amino acid identity, clade-specific signature genes, physiological criteria and ecological habitats. This reclassification reveals the phylogenetic position of the microorganisms and clusters lactobacilli into robust clades with shared ecological and metabolic properties (Zheng et al., 2020).

1.1.6.2 *Lactobacillus delbrueckii*

The species *Lactobacillus delbrueckii* comprises six recognized subspecies, namely *bulgaricus*, *lactis*, *delbrueckii*, *indicus*, *sunkii* and *jakobsenii*. The first three taxa were initially described as distinct species, but they were found to exhibit DNA-DNA homologies of 90-100% among each other and accordingly were united under the same nomenclatural designation (Germond et al., 2003). *Lb. bulgaricus* and *Lb. lactis* present the highest economic importance for the dairy industry since they are widely used in the production of yogurt and various cheese varieties as starter cultures (Chen et al., 2017; Parente et al., 2017). While *Lb. lactis* has been identified mostly in cheeses (Tsakalidou et al., 1994b; Giraffa et al., 2004; El Kafsi et al., 2014b; Hebert et al., 2013), *Lb. bulgaricus* is frequently isolated from fermented milks and yogurt (Tsakalidou et al., 1994b; Song et al, 2016), but it has also been detected in plants (Michaylova et al., 2007). It seems that *Lb. bulgaricus* has rapidly evolved towards its adaptation from a plant-associated niche to the rich in nutrient milk environment, through the loss of redundant traits and protocooperation with *S. thermophilus* (van de Guchte et al., 2006).

Lb. delbrueckii subsp. *delbrueckii*, colonizes vegetable sources and it is unable to ferment lactose and degrade casein, and therefore cannot grow in milk (Weiss et al., 1983; Germond et al., 2003). It has been demonstrated that strains of the subspecies have lost the entire *lac* operon, while *prtB*, although still present, has been mutated by indels at the 3'-end, which may have inactivated the gene (Germond et al., 2003). Bacteriocin- and EPS-producing strains of *Lb. delbrueckii* have been studied *in vitro* for antimicrobial activity against *Klebsiella pneumoniae* (Mogna et al., 2016) and for immunomodulatory properties and antiviral activity against rotavirus in porcine intestinal epithelial cells (Kanmani et al., 2018), respectively. Analysis of *Lb. delbrueckii* EPS gene cluster revealed the presence of *epsA* to *epsE*, *epsK* and *epsN*, which show high percentages of identity with the genes described in *Lb. bulgaricus* (Lamothe et al., 2002), along with seven genes coding for putative glycosyltransferases (Kanmani et al., 2018). Moreover, a number of studies have assessed the probiotic efficacy of the subspecies and the results revealed their positive effects when administered as probiotic dietary supplements to dairy postpartum cows (Otero et al., 2007), sea bass larvae (Silvi et al., 2008) and broilers (Xu, 2008). *Lb. indicus* was isolated from traditional fermented dairy products in India (Dellaglio et al., 2005), *Lb. sunkii* from sunki, a traditional Japanese pickle (Kudo et al., 2012) and *Lb. jakobsenii* from dolo wort, an alcoholic fermented beverage in Burkina Faso (Adimpong et al., 2013). The general genome features of the *Lb. delbrueckii* subspecies strains with complete genomes currently available in the NCBI database are presented in Table 4.

1.1.6.3 *Lactobacillus delbrueckii* subsp. *bulgaricus* and subsp. *lactis*

Lb. bulgaricus is a well known starter for its worldwide application in industrially and naturally fermented dairy products, especially in yogurt, with important strain-dependent technological traits related to the production of flavor compounds (e.g. acetaldehyde and diacetyl), acidifying ability and EPS production (Xu et al., 2015). Multilocus sequence typing (MLST) analysis of seven functional genes of 35 *Lb. bulgaricus* strains related to acetaldehyde production (*ald*, *ldh*, *pdh*, *bgl*, *gpd*, *htp* and *pdc*), revealed significant differences among strains, pinpointing the relationship between phenotypic variability and gene evolution (Liu et al., 2016).

Table 4. General genome features of *Lb. delbrueckii* subspecies strains with complete genomes. Type strains are indicated with a superscript T.

Organism	Strain	Genbank accession	Isolation source	Genome size (bp)	GC (%)	Genes	Proteins	rRNA	tRNA	Pseudogenes	Reference
Lb. bulgaricus	LJJ	NZ_CP049052	Traditional yogurt	1,891,087	49.5	1,961	1,614	24	89	231	Pang et al., 2019
	KLDS1.1011	NZ_CP041280	Traditional dairy	1,887,491	49.8	1,992	1,634	27	95	233	
	MN-BM-F01	NZ_CP013610	Traditional dairy	1,875,071	49.7	1,933	1,585	24	88	233	
	KLDS1.0207	NZ_CP032451	Traditional dairy	1,869,179	49.8	1,974	1,620	27	96	228	
	DSM 20080	NZ_CP019120	Environment	1,868,180	49.8	1,942	1,564	27	95	253	
	ND04	NZ_CP016393	Fermented camel milk	1,861,754	49.6	1,932	1,538	27	95	269	
	ACA-DC 87	NZ_LT899687	Traditional Greek yogurt	1,856,003	49.8	1,928	1,579	27	93	226	Tsakalidou et al., 1994b; Alexandraki et al., 2017b
	L99	NZ_CP017235	Fermented milk	1,848,107	49.7	1,921	1,575	27	95	221	
	2038	NC_017469	-	1,872,918	49.7	1,941	1,562	27	89	260	Zheng et al., 2008; Hao et al., 2011
	ATCC 11842^T	NC_008054	Bulgarian yogurt	1,864,998	49.7	1,940	1,561	27	95	254	Weiss et al., 1983; van de Guchte et al., 2006
	ATCC BAA-365	NC_008529	-	1,856,951	49.7	1,965	1,579	27	98	258	Hao et al., 2011
	ND02	NC_014727	-	2,125,753	49.6	2,133	1,835	27	98	170	Sun et al., 2011a
Lb. lactis	KCCM 34717	NZ_CP018215	Environment	2,263,382	49.1	2,270	1,905	32	98	232	
	KCTC 3034	NZ_CP023139	Sour milk	2,237,608	49.0	2,240	1,889	27	94	227	
	ACA-DC 178	NZ_LS991409	Traditional Greek Kasseri	2,050,316	49.6	2,072	1,694	27	94	254	Tsakalidou et al., 1994b; Alexandraki et al., 2019
	KCTC 3035	NZ_CP018156	-	1,972,735	50.0	1,982	1,697	24	96	162	
	NWC_1_2	CP029250	Natural whey culture from Gruyere cheese	2,250,954	48.6	2,286	1,900	24	94	265	
	DSM 20072^T	NZ_CP022988	Emmental	2,165,984	49.0	2,152	1,800	27	94	228	Weiss et al., 1983
Lb. delbrueckii	NBRC 3202^T	NZ_AP019750	-	1,910,306	50.1	1,918	1,645	24	84	162	Weiss et al., 1983
	TUA4408L	NZ_CP021136	Sunki	2,012,440	49.9	1,999	1,718	27	95	156	
	KCTC 13731	NZ_CP018216	Environment	1,910,506	50.0	1,905	1,600	24	84	194	
Lb. indicus	JCM 15610^T	NZ_CP018614	Environment	1,967,220	49.5	2,000	1,662	27	97	211	Dellaglio et al., 2005
Lb. sunkii	JCM 17838^T	NZ_CP018217	Environment	2,004,337	50.1	1,975	1,726	27	95	124	Kudo et al., 2012
Lb. jakobsenii	DSM 26046^T	NZ_CP018218	Environment	1,891,796	50.1	1,941	1,614	30	127	167	Adimpong et al., 2013

Similarly, a newly developed MLST scheme, using eight housekeeping genes (*clpX, dnaA, groEL, murE, pheS, pyrG, recA* and *rpoB*) was employed for the identification of 251 *Lb. bulgaricus* isolates from a collection of 298 *Lb. delbrueckii* strains. Six lineages were identified amongst the *Lb. bulgaricus* strains, indicating the vast genetic diversity within the subspecies (Song et al., 2016).

EPS production by *Lb. bulgraricus* strains is a desirable technological trait, since it affects significantly the organoleptic characteristics of fermented dairy products, including texture and viscosity. *Lb. bulgaricus* commonly produces heteropolysacharides consisting of repeating units of monomers, such as glucose, galactose and rhamnose (Gruter et al., 1993; Grobben et al., 1997), while it is also able to produce phosphopolysaccharides (Kitazawa et al., 1998). Moreover, the produced EPS are characterized by two different molecular weight EPS fractions (Petry et al., 2003). At the genomic level, the EPS biosynthetic cluster is located on the chromosome, comprising 14 genes (*epsA* to *epsN*), which are organized similarly to the *eps* clusters identified in *S. thermophilus* and *L. lactis* (Jolly and Stingele, 2001; Lamothe et al., 2002). The presence of two distinct EPS gene clusters has also been reported for *Lb. bulgaricus* strains (van de Guchte et al., 2006; Hao et al., 2011).

The probiotic potential of *Lb. bulgaricus* has been extensively studied. Although the adhesion and survival of *Lb. bulgaricus* during passage through the human GIT is ambiguous due to the limited number of bile salt hydrolase genes and the absence of genes encoding mucin-binding proteins (van de Guchte et al., 2006), the regular consumption of yogurt fermented by *Lb. bulgaricus* may facilitate its colonization in the gut (Elli et al., 2006). Antibacterial effect of bacteriocin producing *Lb. bulgaricus* strains against pathogens, such as *Vibrio cholerae*, *Escherichia coli* (Tufail et al., 2011) and *Helicobacter pylori* (Boyanova et al., 2017), have also been reported.

Likewise, *Lb. lactis* constitutes a significant starter, especially for the production of cooked cheeses (Parente et al., 2017). Recently, it was demonstrated that *Lb. lactis* positively influenced both proteolysis and production of volatile compounds during the ripening of raw milk hard cooked cheeses, either through its own enzymatic activities or by providing precursors or nitrogen compounds that favored other species of the cheese microbiota (Buchin et al., 2017).

Genomic analysis corroborates these results, since many components of the proteolytic system (e.g. PrtL, aminopeptidases, endopeptidases, tripeptidase, proline peptidases) have been identified in *Lb. lactis* genome (Hebert et al., 2013).

The inhibitory activity of *Lb. lactis* against food-borne pathogens, such as *L. monocytogenes* (Tharrington and Sorrell, 1992), *E. coli*, *Listeria innocua* (Batdorj et al., 2007), *E. coli* O157:H7 and *Salmonella typhimurium* (Senne and Gilliland, 2003), is attributed mostly to H_2O_2 production. Moreover, strains of *Lb. lactis* have been reported to exert immunomodulatory and anti-inflammatory properties (El Kafsi et al., 2014a; Rolny et al., 2016), as well as to produce health-beneficial bioactive peptides from α_{s1}- and β-casein breakdown (Hebert et al., 2013).

Comparative genomic analysis of *Lb. bulgaricus* and *Lb. lactis* revealed distinct differences between the two subspecies regarding certain genomic features and metabolic traits. As it appears also in Table 4, *Lb. bulgaricus* genomes are smaller than these of *Lb. lactis* (1.8 Mbp and 2 Mbp, respectively), and, as expected, the number of genes encoding proteins is higher in the latter, indicating different metabolic properties between the two subspecies. Of note, El Kafsi et al. (2014b) demonstrated that *Lb. bulgaricus* ND02 strain was initially misclassified and it is a representative strain of *Lb. lactis* rather than *Lb. bulgaricus*.

Although, both *Lb. bulgaricus* and *Lb. lactis* seem to have evolved through reductive processes during adaptation to rich in nutrients dairy niches, the loss of unessential genes is more prominent in *Lb. bulgaricus*. Indeed, *Lb. lactis* has retained more extended carbohydrate metabolizing capabilities than *Lb. bulgaricus*, utilizing a broader repertoire of carbohydrates, including galactose, sucrose, maltose and trehalose, as well as carbohydrates of plant origin (Hebert et al., 2013; El Kafsi et al., 2014b). For example, *in silico* analysis showed that only *Lb. lactis* strains harbor *nplT* and *dexB* implicated in starch degradation. Differences between the two subspecies related to the uptake and metabolism of lactose have been also observed. In contrast to *Lb. bulgaricus*, which excretes the galactose moiety in the medium, *Lb. lactis* has the ability to metabolize galactose, since it owns a lactose PTS system *(lacE/lacF)* for the internalization and phosphorylation of lactose, a phospho-β-galactosidase (*lacG*), which cleaves lactose-6-P, and the required enzymes (*lacAB*, *lacC*, *lacD)* for the metabolism of galactose-6-phosphate. The lactose-galactose antiporter permease (*lacS*) and β-galactosidase (*lacZ*) in *Lb.*

bulgaricus present higher identity to those found in *Streptococcus*, indicating that these genes have most probably been acquired by HGT in the ancestral *Lb. delbrueckii* lineage. Furthermore, amino acid biosynthesis capability is less compromised in *Lb. lactis* in comparison to *Lb. bulgaricus*, since strains of the former are predicted to synthesize Lys, Asp and Arg, while most strains of the latter cannot synthesize these amino acids. Neither of the two subspecies is able to synthesize phenylalanine, tyrosine, tryptophan, Gly, Ser, Ala, Glu, valine, leucine, isoleucine or His. The only significant adaptation of the proteolytic system of both *Lb. bulgaricus* and *Lb. lactis* appears to be the presence of the major cell wall bound protease (PrtB and PrtL, respectively), which is responsible for casein degradation (El Kafsi et al., 2014b; Villegas et al., 2015). This protease has also been found in *Lactobacillus equicursoris*, the closest known relative of *Lb. delbrueckii* (El Kafsi et al., 2014b).

1.1.7 Protocooperation between *Streptococcus thermophilus* and *Lactobacillus bulgaricus*

Yogurt properties are not influenced only by *S. thermophilus* and *Lb. bulgaricus* strain-dependent traits (e.g. EPS production, presence of CEP) but also by the mutually beneficial interactions between the two species, known as protocooperation. This type of bacterial relationship is determined by the production and exchange of growth stimulating metabolites, and results in improved metabolic performance related to enhanced acidification, improved EPS production and affluence of aroma volatiles. The decrease of pH and the production of formic, pyruvic and folic acid, along with CO_2 by *S. thermophilus* promotes the growth of *Lb. bulgaricus*. Subsequently, the latter produces peptides and free amino acids, which stimulate *S. thermophilus* growth (Sieuwerts et al., 2010; Settachaimongkon et al., 2014).

Formic and folic acid and CO_2 produced by *S. thermophilus* are utilized by *Lb. bulgaricus* as precursors or cofactors in purine biosynthesis (Sieuwerts et al., 2010). *S. thermophilus* carries *pflB*, *pflA* and *glyA* genes encoding pyruvate-formate lyase (PFL), pyruvate-formate-lyase activating (PFLA) enzyme and serine hydroxymethyltransferase (SHMT), respectively. The first two are implicated in the production of formic acid while the latter is involved in folate metabolism (Chaves et al., 2003; Nishimura et al., 2013). *Lb. bulgaricus* is unable to produce these metabolites, since is lacking the aforementioned genes (Nishimura et al., 2013). Furthermore, *S.*

thermophilus can generate CO_2 from urea hydrolysis, due to urease activity (encoded by the operon *ureIABCEFGDMQO*; Mora et al., 2004), which affects the optimal growth of *Lb. bulgaricus* during co-cultivation (Driessen et al., 1982). Recently, it was reported that *S. thermophilus* urease activity is vital for yogurt acidification and fermentation acceleration during protocooperation with *Lb. bulgaricus* (Yamauchi et al., 2019).

Concerning the proteolytic ability of the two species, yoghurt starters commonly comprise weakly proteolytic *S. thermophilus* and proteolytic *Lb. bulgaricus* strains. It has been shown that the absence of PrtB adversely affects *S. thermophilus* growth in mixed cultures, while the absence of PrtS does not influence its bacterial growth (Courtin et al., 2002). Furthermore, Settachaimongkon et al. (2014) have demonstrated that only non-proteolytic *S. thermophilus* strains can perform protocooperation with *Lb. bulgaricus*. Although the use of *S. thermophilus* proteolytic strains have been reported to positively affect the acidity of mixed cultures, acceleration of milk fermentation was attributed to the proteolytic activity of *Lb. bulgaricus* strains (Tian et al., 2018). Thus, it seems that proteolysis by *Lb. bulgaricus* is sufficient for the provision of the necessary amino acids to both bacteria. However, transcriptome analysis of mixed *S. thermophilus* and *Lb. bulgaricus* culture revealed upregulation of sulfur and branched-chain amino acids acquisition related-genes. These findings may indicate inadequate proteolytic activity of *Lb. bulgaricus* so as to meet the requirements in these amino acids for both bacteria (Herve-Jimenez et al., 2009; Sieuwerts et al, 2010).

Dissolved oxygen (DO) is another factor that adversely affects yogurt fermentation. It has been demonstrated that the reduction of DO prior to fermentation enhances mainly the growth of *S. thermophilus* and less of *Lb. bulgaricus*, resulting in rapid accumulation of L-lactate and formate, which subsequently stimulate *Lb. bulgaricus* growth. *S. thermophilus* seems to play an active role in DO deduction in milk due to the presence of an H_2O-forming NADH oxidase (encoded by *nox*). Yogurt fermentation with two different pairs of bacterial strains, namely *S. thermophilus* nox^+/*Lb. bulgaricus* and *S. thermophilus* nox^- mutant/*Lb. bulgaricus*, revealed that the latter had diminished fermentation performance, while the DO concentration of the mixed culture remained at high levels even after several hours (Sasaki et al., 2014).

Stress responses during protocooperation have also been studied. H_2O_2 production by *Lb. bulgaricus* facilitates *S. thermophilus* to regulate iron metabolism in order to reduce the production of harmful reactive oxygen species (ROS; Herve-Jimenez et al., 2009; Sieuwerts et al., 2010). In addition, glutathione (GSH) produced by *S. thermophilus* seems to provide protection to both *S. thermophilus* and *Lb. bulgaricus* cells towards acid stress, while the secreted GSH enhances *Lb. bulgaricus* growth (Wang et al., 2016).

EPS production is one of the key factors in yogurt manufacture since it determines the rheology and viscosity of the final product. This trait has been extensively studied primarily in *S. thermophilus* (Cui et al., 2017), but also in *Lb. bulgaricus* (Ali et al., 2019) strains, individually. Sieuwerts et al. (2010) has demonstrated that genes related to EPS production during co-cultivation were upregulated in both microorganisms, in comparison to monocultures. However, the EPS production ability of these bacteria depends on culture conditions, such as temperature and pH (Aslim et al., 2005), as well as the combination of strains employed (Xu et al., 2015).

It has also been demonstrated that the different ratio combinations of *Lb. bulgaricus* to *S. thermophilus* alter the profile of produced volatiles, which define the flavor and aroma of the final product. More specifically, higher concentrations of aldehydes and ketones were identified when the 1:1000 (*Lb. bulgaricus*:*S. thermophilus*) combination of the two species was employed in comparison to other examined combinations (Dan and Wang, 2017). These findings indicate that the selection of both suitable strains and proportions is of great significance for optimum protocooperation of the two species, and, thus, the overall quality of the product (Dan and Wang, 2017).

Finally, this beneficial coexistence during milk fermentation also influences the genetic coevolution of the two species. HGT events have been reported between the two species with regards to EPS biosynthetic genes, transferred from *S. thermophilus* to *Lb. bulgaricus*, and the gene cluster *cbs-cblB(cglB)-cysE* involved in the metabolism of sulfur-containing amino acids, transferred from *Lb. bulgaricus* or *Lb. helveticus* to *S. thermophilus* (Liu et al., 2009).

1.1.8 Bacterial whole-genome sequencing

1.1.8.1 General information

Since the first sequenced bacterial genome (*Haemophilus influenza*; Fleischmann et al., 1995), numerous sequencing techniques have been introduced, rendering whole-genome sequencing (WGS) significantly inexpensive and facile, thus enabling the sequencing of large numbers of bacterial genomes. Indicatively, in 2013 the Genomes Online Database (GOLD; Pagani et al., 2012) listed 2,264 finished bacterial genomes and 4,067 permanent draft genomes (Dark, 2013), whereas nowadays these numbers have been enormously increased, reaching up to 14,758 and 129,264, respectively (accessed on June 2020). Accordingly, 250,377 bacterial genomes have been deposited at the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov, accessed on June 2020), with 18,367 designated as "complete". The majority of these genomes belong to the phyla Proteobacteria (Class Gammaproteobacteria 87,794/6,813 complete), Firmicutes (65,408/4,125 complete) and Actinobacteria (23,723/1,771 complete), indicating the possible socioeconomic importance of several bacterial species included in these phyla.

In addition, the robust combination of genome sequencing and bioinformatics analysis of sequence data has reshaped our perception about bacterial function, evolution and interaction with other microorganisms, with the hosts, and with the ecological niches. Sequence data mining have provided significant insights into microbial diversity (from strains to super-phyla) and have promoted the exploration of microbial communities (Loman and Pallen, 2015).

WGS is increasingly used by food regulatory and public health agencies to facilitate the detection, analysis, and control of foodborne bacterial outbreaks (Brown et al., 2019), in studying and controlling multidrug-resistant (MDR) pathogens (Punina et al., 2015), in microbial forensics for the investigation of bioterrorism (Rasko et al., 2011) and criminal activities (Fierer et al., 2010), in microbial archaelogy (Warinner et al., 2017) and in amplicon and shotgun metagenomics studies (Jovel et al., 2016; Leech et al., 2020). However, the boost in sequencing technologies has led to the accumulation or raw sequence data, shifting the bottleneck in genome sequencing from sequence generation to bioinformatics processing of data (Loman and Pallen, 2015).

1.1.8.2 DNA sequencing technologies

Over the last 50 years, the emergence of different generations of sequencing technology has transformed the landscape of bacterial genome sequencing. Among the first generation sequencing technologies, the major breakthrough concerned the development of Sanger sequencing, which employs a chain-termination approach to DNA sequencing (Sanger et al., 1977). This method is still widely used for sequencing short segments of DNA (up to 1,000 bp) due to its accuracy and ease (Kwong et al., 2015; Heather and Chain, 2016). Subsequently, shotgun sequencing was efficiently used for sequencing longer DNA pieces, even entire chromosomes. The method utilized the random fragmentation of long DNA strands into numerous smaller segments for Sanger sequencing and the produced reads were assembled using their overlaps and/or a genetic map (Staden, 1979). This technique was evolved into whole-genome shotgun sequencing, which uses bacterial clones to produce a large amount of redundant sequence reads across the genome, while the assembly was executed with advanced computational technology and software (Fleischmann et al., 1995).

The second generation or high-throughput sequencing (HTS) technologies are characterized by high accuracy but relatively short read lengths (Loman and Pallen, 2015). Pyrosequencing (454) (Roche Inc., Branford, CT, USA) was introduced in 2005 and was the first next-generation sequencing (NGS) platform on the market (Margulies et al., 2005). This technique uses a “sequencing by synthesis” approach and is based on the measurement of the released inorganic pyrophosphate followed by the incorporation of deoxynucleotides (dNTP), which is converted into visible light during DNA synthesis that is read by the sequencer (Ronaghi et al., 1996; Ronaghi et al., 1998). Pyrosequencing has high error rates (insertions and deletions) in homopolymer repeats because above four or five identical incorporated dNTPs a non-linear readout is produced, meaning that the difference in light intensity between progressively longer nucleotide repeats is relatively less (Margulies et al., 2005; Ronaghi et al., 1998). The first HTS machine widely available to clients was 454 GS 20, superseded later by the 454 GS FLX, which offered a greater number of reads as well as better quality data (Voelkerding et al., 2009).

Sequencing by Oligo Ligation and Detection (SOLiD) system (Life Technologies Corporation, Grand Island, NY, USA) was presented in 2007. This technology was based on a

ligation strategy, relying on the specificity of DNA ligases to ligate fluorescent oligonucleotides to templates in a sequence-dependent manner (Brenner et al., 2000). The most common error type created by SOLiD is substitution, whereas the produced data also revealed an underrepresentation of AT-rich and GC-rich regions (Harismendy et al., 2009). In addition, the method was reported to have some issues in sequencing palindromic sequences (Huang et al., 2012).

Ion Torrent Personal Genome Machine (Life Technologies Corporation, Grand Island, NY, USA) was released in 2011 (Quail et al., 2012). This technology generated 100 bp long reads by measuring the pH variations induced by hydrogen ions released due to nucleotide incorporation during DNA synthesis, using a semiconductor sensor (Rothberg et al., 2011). In 2012 Ion Torrent released the Ion Proton system, which allows for larger chips with more sequence per run. While the method presented is generally of high accuracy, high error rates, particularly in homopolymers repeats, (Loman et al., 2012b) and poor coverage of extremely AT-rich or GC-rich regions (Quail et al., 2012), have been reported.

The Illumina technology (Illumina Inc., San Diego, CA, USA), is based on the Solexa sequencing-by-synthesis method and detects the release of fluorescent labels from incorporated nucleotides for the determination of sequence (Buermans and ded Dunnen, 2014). In 2006, the first Genetic Analyzer GA-II reached the market (Bentley et al., 2008) followed by the moderate throughput platform MiSeq (the only benchtop sequencer, commercialized in 2011), and subsequently by more powerful sequencers, such as NextSeq (moderate-high throughput), HiSeq 2000/2500 (high throughput), TruSeq (long read technology for the production of synthetic reads of 10 kb in length; McCoy et al., 2014), HiSeq X Ten (presented in 2014) and NovaSeq (3.000 Gbp output/run, presented in 2017; Giani et al., 2020). The main drawback of Illumina sequencing systems is the short sequence length (Arredondo-Alonso et al., 2017), which is counterbalanced by the high accuracy, the low error rates and the large amounts of produced raw data (Pallen, 2016). These advantages, along with the limitations of the other sequencing platforms, render the Illumina sequencing technology the most successful nowadays, to the point of near monopoly (Greenleaf and Sidow, 2014), a status that pinpoints the great contribution of the respective technology to the second-generation of DNA sequencers.

The third generation sequencing technologies allow the detection and sequencing of single molecules in real time (SMRT) and their main characteristics comprise low accuracy but long read lengths (Loman and Pallen, 2015). The first single-molecule sequencing technology, based on fluorescence detection and sequencing by synthesis, was Heliscope commercialized in 2009 by Helicos BioSciences (Braslavsky et al., 2003; Harris et al., 2008). Its function was similar to that of Illumina sequencers without the bridge amplification step, but the produced reads were relatively short and the technology was proved to be time and cost consuming. The first approach in SMRT sequencing was commercialized by Pacific Biosciences (PacBio) in 2010. This technology is based on the detection of natural DNA synthesis by a single DNA polymerase molecule (Levene et al., 2003). Incorporation of phosphate-labeled nucleotides leads to base-specific fluorescence, which is detected in real time by the sequencer. The main advantage of PacBio is the long read length. The first instrument, the PacBio RS, appeared in 2010 and generated several thousands of long reads, up to 15 kb (Eid et al., 2009; Loman et al., 2012a). Exceptionally long reads are produced by the new PacBio RS II with maximum read lengths of over 20 kb (Utturkar et al., 2014). The long reads have significantly improved *de novo* assembly and bacterial genome completion without the need of traditional PCR-based gap closure. Although the method undergoes high error rates and the accuracy of individual reads tends to be low (~85-90% base call accuracy in comparison to the 99.9% of Illumina), errors are randomly distributed. Thus, increased coverage allows overlapping reads to produce a consensus sequence with high accuracy (Koren et al., 2012).

In 2019, PacBio released a new method named HiFi (High Fidelity) long-reads, which can generate Circular Consensus Sequences (CCSs) with approximate length from 10 to 20 kb, potentially as accurate as Illumina short reads (Wenger et al., 2019). The same year the company presented a revised platform, namely Sequel II, which is based on new chemicals and it is equipped with a novel 8 million zeromode waveguide (ZMW) nanowells SMRT cell, able to generate a throughput of 160 Gb per SMRT Cell (Giani et al., 2020).

The second approach in SMRT sequencing was Nanopore sequencing, which has been commercialized by Oxford Nanopore Technologies (ONT) since 2005. Nanopore sequencing uses genetically modified bacterial nanopores inserted into an artificial lipid bilayer, placed in

individual microwells arrayed on a sensor chip (Huang et al., 2010; Cherf et al., 2012; Manrao et al., 2012). The nanopores use voltage variation produced when single stranded DNA travels through the channel and the changes induced by each base are measured by a semiconductor sensor and subsequently translated into a DNA sequence. The produced reads are characterized by long length but low accuracy. MinION, GridION, and PromethION platforms are able to sequence very long fragments of DNA (>10 kb, with >2 Mb being reported, the longest DNA reads generated so far; Jain et al., 2018; Payne et al., 2018). Major advantages of nanopore technology are the direct sequencing of DNA or RNA molecules without the need for library preparation or sequencing reagents, the small USB stick-like size of the Nanopore devices and the low cost (van Dijk et al., 2014). However, reads appeared to have sequencing biases difficult to be corrected (Istace et al., 2017). In 2019, Nanopore released R10, a new design of nanopore equipped with a double sensor for improved base calling (Giani et al., 2020). In Table 5 an overview of the different sequencing technologies and their characteristics are presented (Derocles et al., 2018).

Table 5. Comparison of the main sequencing technologies (Derocles et al., 2018)

Sequencing platform	Sequencing method	Read Length (pb)	Error rate (%)	Error Types	Number of reads per run
First Generation					
Sanger ABI 3730xl	Dideoxy chain termination	600-1000	0.001	Insertion Deletion Substitution	96
Second Generation					
454 Roche GS FLX +	Pyrosequencing	700	1	Insertion Deletion	1×10^6
Illumina HiSeq 2500; high output	Synthesis	2 x 125	0.1	Substitution	8×10^9 (paired)
Illumina HiSeq 2500; rapid run	Synthesis	2 x 250	0.1	Substitution	1.2×10^9 (paired)
Illumina MiSeq v3	Synthesis	2 x 300	0.1	Substitution	3×10^8
SOLiD 5500x	Ligation	2 x 60	5	Substitution	8×10^8
Ion Torrent	Polymerase synthesis	200	1	Insertion Deletion	8.2×10^7
Third Generation					
PacBio RS II: P6-C4	Synthesis	~10,000-15,000	13	Insertion Deletion	$3.5\text{-}7.5 \times 10^4$
Oxford Nanopore MinION	Nanopore	~2,000-5,000	38	Insertion Deletion Substitution	$1.7\text{-}4.7 \times 10^4$

1.1.8.3 Sequence assembly

From the biological point of view, the informative value of raw sequence data is increased when these are assembled into long contiguous genomic sequences, either by alignment-based (i.e. against a reference sequence) or assembly-based (i.e. *de novo* assembly) approaches. Downstream applications, such as analysis of genomic variation between species or strains, depend greatly on robust assembled sequences. In this sense, *de novo* assembly of short reads is fundamental to genomics research since it allows the identification of both indels and novel sequence (Loman et al., 2012a), and is particularly useful for sequencing organisms lacking reference sequence.

Several *de novo* genome assemblers using paired-end information for genome scaffolding are widely used in bacterial genome sequencing, with ABySS, SOAPdenovo2, Velvet and SPAdes being among the most popular. ABySS (Assembly By Short Sequences) employs reads produced by first and second generation sequencing technologies (e.g. Illumina, SOLiD, pyrosequencing, Sanger) and implements a distributed representation of de Bruijn graphs, which enable parallel computation of the assembly algorithm (Simpson et al., 2009). Furthermore, the algorithm has been adapted for *de novo* transcriptome assembly using RNA seq data (Trans-ABySS, Robertson et al., 2010). Velvet, one of the first de Bruijn graph-based *de novo* assembler, also utilizes short read sequencing data, similarly to ABySS. The algorithm uses the reads to construct a de Bruijn graph, from which errors are being removed (Zerbino and Birney, 2008). Subsequently, based on the available long reads or paired-end reads, it tries to resolve repeats (Zerbino et al., 2009). In addition, Velvet can be used to analyze colorspace data, i.e. data produced by SOLiD sequencers, although specific settings and use of adjunct conversion programs are required (Zerbino, 2010). Assembly of single-cell data is challenging due to highly non-uniform read coverage as well as multiple sequencing errors and chimeric reads (Bankevich et al., 2012). SPAdes *de novo* genome assembler, which is also based on de Brujin graphs assembly, can generate single-cell assemblies, providing information about genomes of uncultivatable bacteria (Bankevich et al., 2012). SOAPdenovo2 is designed for *de novo* assembly of Illumina reads and comprises multiple modules that execute error correction, assembly, paired end mapping, and scaffold construction (Luo et al., 2012). The assembler is part of a tool package efficiently

implemented in NGS data analysis, namely the Short Oligonucleotide Analysis Package (SOAP), which further includes an alignment tool (SOAPaligner), a re-sequencing consensus sequence builder (SOAPsnp), an indel finder (SOAPindel), a structural variation (SV) scanner, (SOAPsv), and a genome-wide detector of fusion transcripts from paired end RNA-Seq data (SOAPfuse; Gu et al., 2013).

Due to the draft assembly outputs of short-read *de novo* genome assemblers, attention has been given to genome finishing tools, which aim to improve the contiguity of the assembly and to resolve repetitive regions. Single molecule sequencing technologies have facilitated scaffolding of genome assemblies, thus enhancing the quality of incomplete and inaccurate draft assemblies constructed from NGS. For example, SSPACE-LongRead is a hybrid assembly method that aims at scaffold pre-assembled contigs in an iterative manner using PacBio RS long reads as a backbone (Boetzer and Pirovano, 2014). As demonstrated with six bacterial draft genomes assembled using either a single Illumina MiSeq or Roche 454 library, even a 50× coverage of uncorrected PacBio RS long reads was sufficient to reduce the number of contigs, producing bacterial genomes near to completion (Boetzer and Pirovano, 2014). However, a method for *de novo* assembly of data from the Oxford Nanopore MinION, which is able to reconstruct the sequence of an entire bacterial chromosome in a single contig without the need of a reference sequence or data from other sequencing platforms, has also been presented (Loman et al., 2015).

In addition, the availability of multiple long-read assemblers (e.g. Canu, Flye, Miniasm/Minipolish, NECAT, Raven, Redbean, Shasta) is increasing, since long-read sequencing is routinely applied in microbial genomics. Recently, the performance of various long-read assemblers in the context of prokaryote whole genomes was examined, assessing their ability to generate complete assemblies as well as their performance on plasmids and circularization of contigs. Results revealed that none of the examined assemblers constitutes an ideal choice for prokaryote genome long-read assembly, while the overall best performers were Flye (plasmid assembly, best performance at low read depths), Miniasm/Minipolish (clean contig circularization) and Raven (chromosome assembly, high tolerance of low-identity read sets; Wick and Holt, 2019).

In reference-based approaches, the short reads are mapped against a reference sequence (e.g. complete or draft genome, plasmid), and more specifically the reads are assigned to a specific genomic location of the reference (Ronholm et al., 2016). Thus, the choice of an appropriate reference sequence is of great significance. Although, read mapping allows the identification of sequence variants (single nucleotide polymorphism, insertion-deletion) therefore assisting the accurate differentiation between closely related strains (Pightling et al., 2014), it may miss valuable information about novel genomic structures (Loman et al., 2012a). The read length supported by a mapper is an important characteristic (Fonseca et al., 2012). Read mappers such as Bowtie, BWA, GNUMAP, MapReads, Maq, Novoalign, SHRiMP, Stampy and SOAP support short length reads, whereas novel mappers, such as RazerS, BWA-SW, SOAP2, RUM, RMAP, SOAPSplice and Bowtie2, can efficiently manage longer reads (Fonseca et al., 2012). For example, the BLASR (Basic Local Alignment with Successive Refinement) method is used for mapping SMRT sequencing reads with high accuracy and speed, combining data structures from short read alignment with optimization methods from whole genome alignment (Chaisson and Tesler, 2012). Furthermore, read mappers are differentiated by the mode of mapping they employ, to those which attempt to find all mapping locations of each read (e.g. FastHASH, mrsFAST, RazerS3, BitMapper, Hobbes) and to those that use some heuristic methods for identifying one or a few top mapping locations for each read (e.g. Bowtie2, BWA, GEM; Zhang et al., 2018). Fonseca et al. (2012) have reported that several mappers have an option to report all possible mapping locations of a read, but they are less efficient than those specifically designed for this purpose (e.g. mrFast, mrsFAST, PatMaN).

Whole Genome Mapping (WGM; Opgen Inc., Gaithersburg, MD, USA) constitute a useful tool widely used in validation and quality control of assembled genome sequences. This method generates high-resolution restriction maps of a genome, based on the *in situ* restriction digestion of single DNA molecules. The physical genome map provides information about the sizes of the restriction fragments and their physical positions along the DNA strand (Neely et al., 2011). Therefore, it can be used as a reference facilitating not only the accurate arrangement of NGS contigs, but also the identification of gaps along the mapped contigs (Onmus-Leone et al., 2013).

1.1.8.4 Genome annotation

The biological interpretation of assembled genomic sequences is achieved through structural and functional annotation. Structural annotation refers to the identification of various genomic features, such as protein coding sequences, promoters, pseudogenes, RNA genes (e.g. ribosomal RNA genes (rRNAs), transfer RNA genes (tRNAs), small RNAs), and untranslated regions The main difficulty during prediction of coding sequences (CDSs) is the determination of veritable genes from non-coding open reading frames (ORFs; Beckloff et al., 2012).

Gene prediction can be performed either with extrinsic or evidence-based methods (e.g. BLAST comparisons), which depend on *a priori* annotated data stored in biological databases, or with intrinsic or *ab initio* methods, (e.g. Glimmer; Delcher et al., 2007, Prodigal; Hyatt et al., 2010), which use statistical and computational methods for the discrimination between coding and noncoding regions based on DNA sequence properties (e.g. codon, base, amino acid composition; Beckloff et al., 2012). The former approach can be efficiently applied when closely related genomes are available, but not for the annotation of new species or the identification of genes yet to be described. The latter approach has proved to be competent for identification of novel sequences but require very extensive training on large number of high quality gene models. The training set used must be very exact to reach very high accuracy for newly assembled bacterial genomes, which is often hard to accomplish (Beckloff et al., 2012; Wishart, 2020). Furthermore, evidence-based methods, such as GenePRIMP, can be employed for the evaluation and manual curation of gene models in prokaryotic genomes, improving their quality (Pati et al., 2010).

Functional annotation specifies the biological function of the identified DNA-encoded genetic elements, primarily of protein-coding genes, based on sequence similarity to other known genes or proteins (Beckloff et al., 2012). BLASTP suite is the most widely employed tool in similarity searches against various protein sequence databases, such as NCBI non-redundant (nr) database, GenBank Reference Sequence Database (RedSeq), Protein Data Bank (PDB), Universal Protein Resource Knowledgebase (UniProtKB)/Swiss-Prot (Hu and Kurgan, 2018). The resulting sequence alignments are evaluated based on sequence similarity, coverage, E-value and percentage identify cut-offs in relation to the sequences found in the query databases (Beckloff

et al., 2012). Currently, BLASTP suite includes five programs, namely BLASTP, Quick BLASTP, PSI-BLAST (Position-Specific Iterative BLAST), PHI-BLAST (Pattern-Hit Initiated BLAST), and DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST; Hu and Kurgan, 2018).

The utilization of resources that maintain information relative to protein signatures, functional domains, or conserved sites within related protein families, can further facilitate accurate protein annotation (Médigue and Moszer, 2007; Beckloff et al., 2012). These include the Protein Family database (Pfam), which comprises protein domain families based on seed alignments (El-Gebali et al., 2019), the InterPro database, an integrative database of protein families, domains and functional sites that cross-references equivalent entries found in various resources (Apweiler et al., 2000), the Prosite documented database, which consists of a large collection of biologically meaningful motifs linked to documentation briefly describing the protein family or domain they are designed to detect (Sigrist et al., 2002) and NCBI's Conserved Domain Database (CDD) for protein classification (Marchler-Bauer et al., 2005; Marchler-Bauer et al., 2015). The Clusters of Orthologous Groups (COGs) of proteins database has been designed for the classification of proteins from completely sequenced genomes using an orthology-based approach, which confer reliable assignment of orthologs and paralogs for most genes (Tatusov et al., 1997). Comprehensive revision of the COG annotations and inclusion of representative complete genomes from all bacterial and archaeal lineages down to the genus level, has recently been reported (Galperin et al., 2014).

Nowadays, accurate and rapid bacterial genome annotation can be performed with freely available online web services, such as the NCBI prokaryotic genome annotation pipeline (PGAP; Tatusova et al., 2016), and the user-friendly RAST Server (Overbeek et al., 2014). Annotated genome files can be further visually explored for genomic features of interest, compared to other genomes and manually curated using a variety of tools, such as the Artemis genome browser (Carver et al., 2012) and the BLAST suite (Altschul et al., 1990).

The biological role of annotated genes and proteins in an organism can be further elucidated through the reconstruction of metabolic pathways they are involved in. This can be achieved by employing the gene ontology (GO) system, which assigns basic functionality to a hypothetical gene/protein or more specific descriptions to those already described, as well as

accurate annotation of orthologs (Ashburner et al., 2000; Gene Ontology Consortium, 2015), complemented by the widely popular Kyoto Encyclopedia of Genes and Genomes (KEGG), a knowledgebase of genomic and molecular information (Kanehisa and Goto, 2000). KEGG currently includes three main databases, namely PATHWAY, BRITE, and MODULE, which contain experimental knowledge of high-level functions from published literature, represented in terms of KEGG pathway maps, BRITE hierarchies, and KEGG modules, respectively (Kanehisa and Sato, 2020). The pathway mapping tools included in KEGG Mapper have been exploited in biological data mining from numerous genome sequences (Kanehisa and Sato, 2020).

Functional annotation is often challenging due to evolutionary processes. Gene structure similarity does not necessarily ensure function consistency and vice versa, while gene duplication can result in the formation of paralogs. Genes can gain or lose functional domains and fuse with adjacent genes, causing the formation of chimeric proteins that share multiple ancestors (Beckloff et al., 2012). Comparative genome analysis aims at ascertaining biological relevance from the observed similarities and differences among genomes and can be efficiently used for the identification of mis-annotation in microbial genomes (Poptsova and Gogarten, 2010). Several comparative genomics resources have been developed (e.g. progressiveMauve; Darling et al., 2010, the Integrated Microbial Genomes (IMG) system; Markowitz et al., 2011, the microbial genome database (MBGD) for comparative analysis; Uchiyama, 2006, MicroScope; Vallenet et al., 2009; Vallenet et al., 2020), to support and facilitate the concurrent analysis of multiple microbial genomes.

1.2 Aim of the thesis

LAB are indisputably among the most technologically significant groups of bacteria since their diverse functional traits are being applied in various industrial sectors, such as food and metabolite production, health promotion, animal husbandry and agricultural systems. These functions, encoded within genomes, have been studied extensively over the years due to the advance of both sequencing technologies and bioinformatics methods. Analysis of strain-dependent molecular pathways has resulted in more accurate use of LAB in commercial applications, such as starter/adjunct cultures and probiotic supplements. Thus, the genomic

decryption of subjacent biological qualities of various LAB strains enhances our knowledge concerning their role in current or even novel applications.

Furthermore, the isolation on new LAB strains from traditionally fermented products, which could be used as starters or adjuncts in industrial fermentations, is a common practice. The ACA-DC collection of the Laboratory of Dairy Research of the Agricultural University of Athens is a repository of such "wild" microorganisms. The collection currently hosts a great number of bacterial strains isolated from naturally fermented Greek traditional foods, the majority of which are yet to be studied in detail at the genome level.

Therefore, the aim of the present thesis was to decipher the biological and technological potential of three LAB strains of the ACA-DC collection, namely *S. thermophilus* ACA-DC 2, *Lb. bulgaricus* ACA-DC 87 and *Lb. lactis* ACA-DC 178*,* isolated from traditional Greek dairy products, namely yogurt and Kasseri cheese, employing NGS technologies and efficient bioinformatics tools. More specifically, validation of the genome assembly, structural and functional annotation of the final chromosomal sequence as well as comparative genomics against closely related strains and/or species, was performed for each of the respective strains. Comparative and evolutionary genomic analysis was focused mainly and performed in detail for *S. thermophilus* species, employing 23 *S. thermophilus* strains with publicly available complete genomes, including strain ACA-DC 2. The analysis aimed at interpreting the clonal relations among the strains through the identification and evaluation of presence/absence patterns of certain genomic traits as well as in pinpointing strain-specific differences.

1.3 References

Abe, K., Ohnishi, F., Yagi, K., Nakajima, T., Higuchi, T., Sano, et al. (2002). Plasmid-encoded asp operon confers a proton motive metabolic cycle catalyzed by an aspartate-alanine exchange reaction. *J. Bacteriol.* 184**,** 2906. doi: 10.1128/JB.184.11.2906-2913.2002.

Adimpong, D.B., Nielsen, D.S., Sorensen, K.I., Vogensen, F.K., Sawadogo-Lingani, H., Derkx, P.M.F., et al. (2013). *Lactobacillus delbrueckii* subsp. *jakobsenii* subsp. nov., isolated from dolo wort, an alcoholic fermented beverage in Burkina Faso. *Int. J. Syst. Evol. Microbiol.* 63**,** 3720-3726. doi: 10.1099/ijs.0.048769-0.

Adu, K.T., Wilson, R., Nichols, D.S., Baker, A.L., Bowman, J.P., and Britz, M.L. (2018). Proteomic analysis of *Lactobacillus casei* GCRL163 cell-free extracts reveals a SecB homolog and other biomarkers of prolonged heat stress. *PloS One* 13, e0206317-e0206317. doi: 10.1371/journal.pone.0206317.

Afzal, M.I., Jacquet, T., Delaunay, S., Borges, F., Millière, J.-B., Revol-Junelles, A.-M., et al. (2010). *Carnobacterium maltaromaticum*: Identification, isolation tools, ecology and technological aspects in dairy products. *Food Microbiol.* 27**,** 573-579. doi: 10.1016/j.fm.2010.03.019.

Al-Naseri, A., Bowman, J.P., Wilson, R., Nilsson, R.E., and Britz, M.L. (2013). Impact of lactose starvation on the physiology of *Lactobacillus casei* GCRL163 in the presence or absence of tween 80. *J. Proteome Res.* 12**,** 5313-5322. doi: 10.1021/pr400661g.

Alewijn, M., Sliwinski, E.L., and Wouters, J.T.M. (2005). Production of fat-derived (flavour) compounds during the ripening of Gouda cheese. *Int. Dairy J.* 15**,** 733-740. doi: 10.1016/j.idairyj.2004.09.009.

Alewijn, M., Smit, B.A., Sliwinski, E.L., and Wouters, J.T.M. (2007). The formation mechanism of lactones in Gouda cheese. *Int. Dairy J.* 17**,** 59-66. doi: 10.1016/j.idairyj.2006.01.002.

Alexandraki, V., Kazou, M., Blom, J., Pot, B., Tsakalidou, E., and Papadimitriou, K. (2017a). The complete genome sequence of the yogurt isolate *Streptococcus thermophilus* ACA-DC 2. *Stand. Genomic Sci.* 12:18. doi: 10.1186/s40793-017-0227-5

Alexandraki, V., Kazou, M., Pot, B., Tsakalidou, E., and Papadimitriou, K. (2017b). Complete genome sequence of the yogurt isolate *Lactobacillus delbrueckii* subsp. bulgaricus ACA-DC 87. *Genome Announc.* 5**,** e00868-00817. doi: 10.1128/genomeA.00868-17.

Alexandraki, V., Kazou, M., Pot, B., Tsakalidou, E., and Papadimitriou, K. (2019). Whole-genome sequence data and analysis of *Lactobacillus delbrueckii* subsp. *lactis* ACA-DC 178 isolated from Greek Kasseri cheese. *Data Brief* 25**,** 104282. doi: 10.1016/j.dib.2019.104282.

Ali, K., Mehmood, M.H., Iqbal, M.A., Masud, T., Qazalbash, M., Saleem, S., et al. (2019). Isolation and characterization of exopolysaccharide-producing strains of *Lactobacillus bulgaricus* from curd. *Food Sci. Nutr.* 7**,** 1207-1213. doi: 10.1002/fsn3.905.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215**,** 403-410. doi: 10.1016/s0022-2836(05)80360-2.

Alvarez-Sieiro, P., Montalban-Lopez, M., Mu, D., and Kuipers, O.P. (2016). Bacteriocins of lactic acid bacteria: Extending the family. *Appl. Microbiol. Biotechnol.* 100**,** 2939-2951. doi: 10.1007/s00253-016-7343-9.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., et al. (2000). InterPro-an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics* 16**,** 1145-1150. doi: 10.1093/bioinformatics/16.12.1145.

Araque, I., Gil, J., Carreté, R., Bordons, A., and Reguant, C. (2009). Detection of arc genes related with the ethyl carbamate precursors in wine lactic acid bacteria. *J. Agric. Food Chem.* 57**,** 1841-1847. doi: 10.1021/jf803421w.

Ardo, Y. (2006). Flavour formation by amino acid catabolism. *Biotechnol. Adv.* 24**,** 238-242. doi: 10.1016/j.biotechadv.2005.11.005.

Arredondo-Alonso, S., Willems, R.J., Van Schaik, W., and Schürch, A.C. (2017). On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb. Genom.* 3**,** e000128-e000128. doi: 10.1099/mgen.0.000128.

Arioli, S., Monnet, C., Guglielmetti, S., Parini, C., De Noni, I., Hogenboom, J., et al. (2007). Aspartate biosynthesis is essential for the growth of *Streptococcus thermophilus* in milk, and aspartate

availability modulates the level of urease activity. *Appl. Environ. Microbiol.* 73, 5789–5796. doi: 10.1128/AEM.00533-07.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., et al. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics* 25**,** 25-29. doi: 10.1038/75556.

Aslim, B., Yüksekdag, Z.N., Beyatli, Y., and Mercan, N. (2005). Exopolysaccharide production by *Lactobacillus delbruckii* subsp. *bulgaricus* and *Streptococcus thermophilus* strains under different growth conditions. *World J. Microbiol. Biotechnol.* 21**,** 673-677. doi: 10.1007/s11274-004-3613-2.

Bai, Y., Sun, E., Shi, Y., Jiang, Y., Chen, Y., Liu, S., et al. (2016). Complete genome sequence of *Streptococcus thermophilus* MN-BM-A01, a strain with high exopolysaccharides production. *J. Biotechnol.* 224, 45–46. doi: 10.1016/j.jbiotec.2016.03.003.

Bailey, J.R., Vince, V., Williams, N.A., and Cogan, T.A. (2017). *Streptococcus thermophilus* NCIMB 41856 ameliorates signs of colitis in an animal model of inflammatory bowel disease. *Benef. Microbes* 8**,** 605-614. doi: 10.3920/BM2016.0110.

Baliarda, A., Robert, H., Jebbar, M., Blanco, C., Deschamps, A., and Le Marrec, C. (2003). Potential osmoprotectants for the lactic acid bacteria *Pediococcus pentosaceus* and *Tetragenococcus halophila*. *Int. J. Food Microbiol.* 84**,** 13-20. doi: 10.1016/s0168-1605(02)00388-4.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19**,** 455-477. doi: 10.1089/cmb.2012.0021.

Barbieri, F., Montanari, C., Gardini, F., and Tabanelli, G. (2019). Biogenic amine production by lactic acid bacteria: A review. *Foods* 8**,** 17. doi: 10.3390/foods8010017.

Barbieri, G., Bolzoni, L., Careri, M., Mangia, A., Parolari, G., Spagnoli, S., et al. (1994). Study of the volatile fraction of parmesan cheese. *J. Agric. Food Chem.* 42**,** 1170-1176. doi: 10.1021/jf00041a023.

Barrangou, R., and Horvath, P. (2012). CRISPR: New horizons in phage resistance and strain identification. *Annu. Rev. of Food Sci. Technol.* 3**,** 143-162. doi: 10.1146/annurev-food-022811-101134.

Batdorj, B., Trinetta, V., Dalgalarrondo, M., Prévost, H., Dousset, X., Ivanova, I., et al. (2007). Isolation, taxonomic identification and hydrogen peroxide production by *Lactobacillus delbrueckii* subsp. *lactis* T31, isolated from Mongolian yoghurt: Inhibitory activity on food-borne pathogens. *J. Appl. Microbiol.* 103**,** 584-593. doi: 10.1111/j.1365-2672.2007.03279.x.

Beckloff, N., Starkenburg, S., Freitas, T., and Chain, P. (2012). "Bacterial genome annotation," in *Microbial Systems Biology, Methods in Molecular Biology (Methods and Protocols)*, ed. A. Navid (Totowa, NJ: Humana Press), 471-503. doi: 10.1007/978-1-61779-827-6_16.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456**,** 53-59. doi: 10.1038/nature07517.

Bentley, R.W., Leigh, J.A., and Collins, M.D. (1991). Intrageneric structure of *Streptococcus* based on comparative analysis of small-subunit rRNA sequences. *Int. J. Syst. Bacteriol.* 41**,** 487-494. doi: 10.1099/00207713-41-4-487.

Beshkova, D.M., Simova, E.D., Simov, Z.I., Frengova, G.I., and Spasov, Z.N. (2002). Pure cultures for making kefir. *Food Microbiol.* 19**,** 537-544. doi: 10.1006/fmic.2002.0499.

Biavati, B. (2001). "Bifidobacteria," in *Microorganisms As Health Supporters*, eds B. Biavati, V. Bottazzi, L. Morelli & C. Schiavi (Novara: Mofin-Alce), 10-33.

Blaya, J., Barzideh, Z., and Lapointe, G. (2018). Symposium review: Interaction of starter cultures and nonstarter lactic acid bacteria in the cheese environment. *J. Dairy Sci.* 101**,** 3611-3629. doi: 10.3168/jds.2017-13345.

Boels, I.C., Van Kranenburg, R., Kanning, M.W., Chong, B.F., De Vos, W.M., and Kleerebezem, M. (2003). Increased exopolysaccharide production in *Lactococcus lactis* due to increased levels of expression of the NIZO B40 eps gene cluster. *Appl. Environ. Microbiol.* 69**,** 5029-5031. doi: 10.1128/aem.69.8.5029-5031.2003.

Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinform.* 15**,** 211. doi: 10.1186/1471-2105-15-211.

Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S. D., Kulakauskas, S., et al. (2004). Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat. Biotechnol.* 22, 1554–1558. doi: 10.1038/nbt1034

Börner, R.A., Kandasamy, V., Axelsen, A.M., Nielsen, A.T., and Bosma, E.F. (2019). Genome editing of lactic acid bacteria: Opportunities for food, feed, pharma and biotech. *FEMS Microbiol. Lett.* 366. doi: 10.1093/femsle/fny291.

Bourdichon, F., Casaregola, S., Farrokh, C., Frisvad, J.C., Gerds, M.L., Hammes, W.P., et al. (2012). Food fermentations: Microorganisms with technological beneficial use. *Int. J. Food Microbiol.* 154**,** 87-97. doi: 10.1016/j.ijfoodmicro.2011.12.030.

Boyanova, L., Gergova, G., Markovska, R., Yordanov, D., and Mitov, I. (2017). Bacteriocin-like inhibitory activities of seven *Lactobacillus delbrueckii* subsp. *bulgaricus* strains against antibiotic susceptible and resistant *Helicobacter pylori* strains. *Lett. Appl. Microbiol.* 65**,** 469-474. doi: 10.1111/lam.12807.

Braslavsky, I., Hebert, B., Kartalov, E., and Quake, S.R. (2003). Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* 100**,** 3960-3964. doi: 10.1073/pnas.0230489100.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* 18**,** 630-634. doi: 10.1038/76469.

Broadbent, J.R., Houck, K., Johnson, M.E., and Oberg, C.J. (2003a). Influence of adjunct use and cheese microenvironment on nonstarter bacteria in reduced-fat cheddar-type cheese. *J. Dairy Sci.* 86**,** 2773-2782. doi: 10.3168/jds.S0022-0302(03)73874-0.

Broadbent, J.R., Mcmahon, D.J., Welker, D.L., Oberg, C.J., and Moineau, S. (2003b). Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: A review. *J. Dairy Sci.* 86**,** 407-423. doi: 10.3168/jds.S0022-0302(03)73619-4.

Brown, E., Dessai, U., Mcgarry, S., and Gerner-Smidt, P. (2019). Use of whole-genome sequencing for food safety and public health in the United States. *Foodborne Pathog. Dis.* 16**,** 441-450. doi: 10.1089/fpd.2019.2662.

Buchin, S., Duboz, G., and Salmon, J.-C. (2017). *Lactobacillus delbrueckii* subsp. *lactis* as a starter culture significantly affects the dynamics of volatile compound profiles of hard cooked cheeses. *Eur. Food Res. Technol.* 243**,** 1943-1955. doi: 10.1007/s00217-017-2899-x.

Buermans, H.P.J., and Den Dunnen, J.T. (2014). Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta Mol. Basis Dis.* 1842**,** 1932-1941. doi: 10.1016/j.bbadis.2014.06.015.

Calles-Enríquez, M., Eriksen, B.H., Andersen, P.S., Rattray, F.P., Johansen, A.H., Fernández, M., et al. (2010). Sequencing and transcriptional analysis of the *Streptococcus thermophilus* histamine biosynthesis gene cluster: factors that affect differential hdcA expression. *Appl. Environ. Microbiol.* 76**,** 6231-6238. doi: 10.1128/AEM.00827-10.

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., and Mcquillan, J.A. (2012). Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28**,** 464-469. doi: 10.1093/bioinformatics/btr703.

Castillo, I., Requena, T., De Palencia, P.F., Fontecha, J., and Gobbetti, M. (1999). Isolation and characterization of an intracellular esterase from *Lactobacillus casei* subsp. *casei* IFPL731. *J. Appl. Microbiol.* 86**,** 653-659. doi: 10.1046/j.1365-2672.1999.00708.x.

Chaisson, M.J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinform.* 13**,** 238. doi: 10.1186/1471-2105-13-238.

Chaves, A.C.S.D., Ruas-Madiedo, P., Starrenburg, M., Hugenholtz, J., and Lerayer, A.L.S. (2003). Impact of engineered *Streptococcus thermophilus* strains overexpressing *glyA* gene on folic acid and acetaldehyde production in fermented milk. *Braz. J. Microbiol.* 34, 114-117.

Chen, Y., and Zhang, M. (2019). A role of exopolysaccharide produced by *Streptococcus thermophilus* in the intestinal inflammation and mucosal barrier in Caco-2 monolayer and dextran sulphate sodium-induced experimental murine colitis. *Molecules* 24, 513. doi: 10.3390/molecules24030513.

Chen, C., Zhao, S., Hao, G., Yu, H., Tian, H., and Zhao, G. (2017). Role of lactic acid bacteria on the yogurt flavour: A review. *Int. J. Food Prop.* 20**,** S316-S330. doi: 10.1080/10942912.2017.1295988.

Cherf, G.M., Lieberman, K.R., Rashid, H., Lam, C.E., Karplus, K., and Akeson, M. (2012). Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.* 30**,** 344-348. doi: 10.1038/nbt.2147.

Chich, J.-F., Marchesseau, K., and Gripon, J.-C. (1997). Intracellular esterase from *Lactococcus lactis* subsp. *lactis* NCDO 763: Purification and characterization. *Int. Dairy J.* 7**,** 169-174. doi: 10.1016/S0958-6946(97)00001-0.

Chun, L., Li-Bo, L., Di, S., Jing, C., and Ning, L. (2012). Response of osmotic adjustment of *Lactobacillus bulgaricus* to NaCl Stress. *J. Northeast Agric. Univ.* 19**,** 66-74. doi: 10.1016/S1006-8104(13)60054-9.

Cogan, T., Barbosa, M., Beuvier, E., Bianchi-Salvadori , B., Cocconcelli, P.S., Fernandes , I., et al. (1997). Characterization of lactic acid bacteria in artisanal dairy products. *J. Dairy Res.* 64**,** 409-421. doi: 10.1017/S0022029997002185.

Cole, J.N., Henningham, A., Gillen, C.M., Ramachandran, V., and Walker, M.J. (2008). Human pathogenic streptococcal proteomics and vaccine development. *Proteomics Clin. Appl.* 2**,** 387-410. doi: 10.1002/prca.200780048.

Collins, Y.F., Mcsweeney, P.L.H., and Wilkinson, M.G. (2003). Lipolysis and free fatty acid catabolism in cheese: A review of current knowledge. *Int. Dairy J.* 13**,** 841-866. doi: 10.1016/S0958-6946(03)00109-2.

Courtin, P., Monnet, V., and Rul, F. (2002). Cell-wall proteinases PrtS and PrtB have a different role in *Streptococcus thermophilus*/*Lactobacillus bulgaricus* mixed cultures in milk. *Microbiology* 148, 3413–3421. doi: 10.1099/00221287-148-11-3413.

Crawley, A.B., Henriksen, E.D., Stout, E., Brandt, K., and Barrangou, R. (2018). Characterizing the activity of abundant, diverse and active CRISPR-Cas systems in lactobacilli. *Sci. Rep.* 8**,** 11544. doi: 10.1038/s41598-018-29746-3.

Cui, Y., Jiang, X., Hao, M., Qu, X., and Hu, T. (2017). New advances in exopolysaccharides production of *Streptococcus thermophilus*. *Arch. Microbiol.* 199**,** 799-809. doi: 10.1007/s00203-017-1366-1.

Cui, Y., Xu, T., Qu, X., Hu, T., Jiang, X., and Zhao, C. (2016). New insights into various production characteristics of *Streptococcus thermophilus* strains. *Int. J. Mol. Sci.* 17. doi: 10.3390/ijms17101701.

Dabour, N., and Lapointe, G. (2005). Identification and molecular characterization of the chromosomal exopolysaccharide biosynthesis gene cluster from *Lactococcus lactis* subsp. *cremoris* SMQ-461. *Appl. Environ. Microbiol.* 71**,** 7414-7425. doi: 10.1128/AEM.71.11.7414-7425.2005.

Dan, T., and Wang, D. (2017). Profiles of volatile flavor compounds in milk fermented with different proportional combinations of *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus*. *Molecules* 22, 1633. doi: 10.3390/molecules22101633.

Dark, M.J. (2013). Whole-genome sequencing in bacteriology: State of the art. *Infect. Drug Resist.* 6**,** 115-123. doi: 10.2147/IDR.S35710.

Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5**,** e11147. doi: 10.1371/journal.pone.0011147.

De Angelis, M., and Gobetti, M. (2011). "Stress responses of lactobacilli," in *Stress Responses of Lactic Acid Bacteria*, eds E. Tsakalidou, and K. Papadimitriou, (New York, NY: Springer ScienceCBusiness Media), 219-249. doi: 10.1007/978-0-387-92771-8_11

De Angelis, M., Di Cagno, R., Huet, C., Crecchio, C., Fox, P.F., and Gobbetti, M. (2004). Heat shock response in *Lactobacillus plantarum*. *Appl. Environ. Microbiol.* 70, 1336-1346. doi: 10.1128/aem.70.3.1336-1346.2004.

De Vin, F., Rådström, P., Herman, L., and De Vuyst, L. (2005). Molecular and biochemical analysis of the galactose phenotype of dairy *Streptococcus thermophilus* strains reveals four different fermentation profiles. *Appl. Environ. Microbiol.* 71**,** 3659. doi: 10.1128/AEM.71.7.3659-3667.2005.

De Vuyst, L., and Leroy, F. (2007). Bacteriocins from lactic acid bacteria: Production, purification, and food applications. *J. Mol. Microbiol. Biotechnol.* 13**,** 194-199. doi: 10.1159/000104752.

De Vuyst, L., Schrijvers, V., Paramithiotis, S., Hoste, B., Vancanneyt, M., Swings, J., et al. (2002). The biodiversity of lactic acid bacteria in Greek traditional wheat sourdoughs is reflected in both

composition and metabolite formation. *Appl. Environ. Microbiol.* 68**,** 6059-6069. doi: 10.1128/aem.68.12.6059-6069.2002.

Delcher, A.L., Bratke, K.A., Powers, E.C., and Salzberg, S.L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23**,** 673-679. doi: 10.1093/bioinformatics/btm009.

Dellaglio, F., Felis, G.E., Castioni, A., Torriani, S., and Germond, J.E. (2005). *Lactobacillus delbrueckii* subsp. *indicus* subsp. nov., isolated from Indian dairy products. *Int. J. Syst. Evol. Microbiol.* 55**,** 401-404. doi: 10.1099/ijs.0.63067-0.

Delorme, C., Abraham, A.L., Renault, P., and Guédon, E. (2015). Genomics of *Streptococcus salivarius*, a major human commensal. *Infect. Genet. Evol.* 33**,** 381-392. doi: 10.1016/j.meegid.2014.10.001.

Delorme, C., Bartholini, C., Luraschi, M., Pons, N., Loux, V., Almeida, M., et al. (2011). Complete genome sequence of the pigmented *Streptococcus thermophilus* strain JIM8232. *J. Bacteriol.* 193**,** 5581-5582. doi: 10.1128/jb.05404-11.

Delorme, C., Legravet, N., Jamet, E., Hoarau, C., Alexandre, B., El-Sharoud, W.M., et al. (2017). Study of *Streptococcus thermophilus* population on a world-wide and historical collection by a new MLST scheme. *Int. J. Food Microbiol.* 242**,** 70-81. doi: 10.1016/j.ijfoodmicro.2016.11.016.

Derocles, S.A.P., Bohan, D.A., Dumbrell, A.J., Kitson, J.J.N., Massol, F., Pauvert, C., et al. (2018). "Biomonitoring for the 21st century: Integrating next-generation sequencing into ecological network analysis," in *Advances in Ecological Research,* eds. D.A. Bohan, A.J. Dumbrell, G. Woodward & M. Jackson. (Academic Press), 1-62.

Dertli, E., Colquhoun, I.J., Gunning, A.P., Bongaerts, R.J., Le Gall, G., Bonev, B.B., et al. (2013). Structure and biosynthesis of two exopolysaccharides produced by *Lactobacillus johnsonii* FI9785. *J. Biol. Chem.* 288**,** 31938-31951. doi: 10.1074/jbc.M113.507418.

Derzelle, S., Hallet, B., Ferain, T., Delcour, J., and Hols, P. (2003). Improved adaptation to cold-shock, stationary-phase, and freezing stresses in *Lactobacillus plantarum* overproducing cold-shock proteins. *Appl. Environ. Microbiol.* 69**,** 4285-4290. doi: 10.1128/aem.69.7.4285-4290.2003.

Desfossés-Foucault, É., Lapointe, G., and Roy, D. (2014). Transcription profiling of interactions between *Lactococcus lactis* subsp. *cremoris* SK11 and *Lactobacillus paracasei* ATCC 334 during Cheddar cheese simulation. *Int. J. Food Microbiol.* 178**,** 76-86. doi: 10.1016/j.ijfoodmicro.2014.03.004.

Desmond, C., Fitzgerald, G.F., Stanton, C., and Ross, R.P. (2004). Improved stress tolerance of GroESL-overproducing *Lactococcus lactis* and probiotic *Lactobacillus paracasei* NFBC 338. *Appl. Environ. Microbiol.* 70, 5929-5936. doi: 10.1128/aem.70.10.5929-5936.2004.

Dimopoulou, M., Vuillemin, M., Campbell-Sills, H., Lucas, P.M., Ballestra, P., Miot-Sertier, C., et al. (2014). Exopolysaccharide (EPS) synthesis by *Oenococcus oeni*: From genes to phenotypes. *PloS One* 9**,** e98898-e98898. doi: 10.1371/journal.pone.0098898.

Dirinck, P., and De Winne, A. (1999). Flavour characterisation and classification of cheeses by gas chromatographic–mass spectrometric profiling. *J. Chromatogr. A* 847**,** 203-208. doi: 10.1016/S0021-9673(99)00193-4.

Divya, J.B., Varsha, K.K., Nampoothiri, K.M., Ismail, B., and Pandey, A. (2012). Probiotic fermented foods for health benefits. *Eng. Life Sci.* 12. doi: 10.1002/elsc.201100179.

Donot, F., Fontana, A., Baccou, J.C., and Schorr-Galindo, S. (2012). Microbial exopolysaccharides: Main examples of synthesis, excretion, genetics and extraction. *Carbohydr. Polym.* 87**,** 951-962. doi: 10.1016/j.carbpol.2011.08.083.

Doroshchuk, N.A., Gelfand, M.S., and Rodionov, D.A. (2006). Regulation of nitrogen metabolism in gram-positive bacteria. *Mol. Biol.* 40**,** 829-836. doi: 10.1134/S0026893306050190.

Drider, D., Bekal, S., and Prevost, H. (2004). Genetic organization and expression of citrate permease in lactic acid bacteria. *Genet. Mol. Res.* 3**,** 273-281.

Driessen, F.M., Kingma, F., Stadhouders, J., 1982. Evidence that *Lactobacillus bulgaricus* in yogurt is stimulated by carbon dioxide produced by *Streptococcus thermophilus*. *Neth. Milk Dairy J.* 36, 135 -144.

Drouault, S., Anba, J., and Corthier, G. (2002). *Streptococcus thermophilus* is able to produce a beta-galactosidase active during its transit in the digestive tract of germ-free mice. *Appl. Environ. Microbiol.* 68**,** 938-941. doi: 10.1128/aem.68.2.938-941.2002.

Dupuis, M.-È., Villion, M., Magadán, A.H., and Moineau, S. (2013). CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat. Commun.* 4:2087. doi: 10.1038/ncomms3087.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA sequencing from single polymerase molecules. *Science* 323**,** 133. doi: 10.1126/science.1162986.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47**,** D427-D432. doi: 10.1093/nar/gky995.

El Kafsi, H., Binesse, J., Loux, V., Buratti, J., Boudebbouze, S., Dervyn, R., et al. (2014a). Genome sequence of *Lactobacillus delbrueckii* subsp. *lactis* CNRZ327, a dairy bacterium with anti-inflammatory properties. *Genome Announc.* 2**,** e00328-00314. doi: 10.1128/genomeA.00328-14.

El Kafsi, H., Binesse, J., Loux, V., Buratti, J., Boudebbouze, S., Dervyn, R., et al. (2014b). *Lactobacillus delbrueckii* ssp. lactis and ssp. *bulgaricus*: A chronicle of evolution in action. *BMC Genom.* 15**,** 407. doi: 10.1186/1471-2164-15-407.

El Soda, M., Howaida Abd El Wahab, H., Ezzat, N., Desmazeaud, M.J., and Ismail, A. (1986) The esterolytic and lipolytic activities of the Lactobacilli. II. Detection of the esterase system of *Lactobacillus helveticus*, *Lactobacillus bulgaricus*. *Lactobacillus lactis* and *Lactobacillus acidophilus*. *Le Lait*, 66 (4), 431-443.

Elli, M., Callegari, M.L., Ferrari, S., Bessi, E., Cattivelli, D., Soldi, S., et al. (2006). Survival of yogurt bacteria in the human gut. *Appl. Environ. Microbiol.* 72**,** 5113-5117. doi: 10.1128/AEM.02950-05.

Eng, C., Thibessard, A., Danielsen, M., Rasmussen, T.B., Mari, J. F., and Leblond, P. (2011). *In silico* prediction of horizontal gene transfer in *Streptococcus thermophilus*. *Arch. Microbiol.* 193, 287–297. doi: 10.1007/s00203-010-0671-8.

Engels, W.J.M., Dekker, R., De Jong, C., Neeter, R., and Visser, S. (1997). A comparative study of volatile compounds in the water-soluble fraction of various types of ripened cheese. *Int. Dairy J.* 7**,** 255-263. doi: 10.1016/S0958-6946(97)00003-4.

Esteban-Torres, M.A., Blanca De Las, R., José Miguel, M.O., and Rosario, M.O. (2015). Characterization of a halotolerant lipase from the lactic acid bacteria *Lactobacillus plantarum* useful in food fermentations. *LWT-Food Sci. Technol.* 60**,** 246-252. doi: 10.1016/j.lwt.2014.05.063.

European Food Safety Authority [EFSA], (2007). Opinion of the scientific committee on a request from EFSA on the introduction of a Qualified Presumption of Safety (QPS) approach for assessment of selected microorganisms referred to EFSA. *EFSA J.* 587, 1-16. doi: 10.2903/j.efsa.2007.587.

Evivie, S.E., Li, B., Ding, X., Meng, Y., Yu, S., Du, J., et al. (2017). Complete genome sequence of *Streptococcus thermophilus* KLDS 3.1003, a strain with high antimicrobial potential against foodborne and vaginal pathogens. *Front. Microbiol.* 8**,** 1238-1238. doi: 10.3389/fmicb.2017.01238.

FAO/WHO (2001) Probiotics in food: Health and nutritional properties and guidelines for evaluation. *Report of a joint FAO/WHO expert consultation on evaluation of health and nutritional properties of probiotics in food including powder milk with live lactic acid bacteria*. Cordoba, AR: FAO/WHO.

Fenster, K.M., Parkin, K., and Steele, J. (2003a). Intracellular esterase from *Lactobacillus casei* LILA: Nucleotide sequencing, purification, and characterization. *J. Dairy Sci.* 86**,** 1118-1129. doi: 10.3168/jds.S0022-0302(03)73694-7.

Fenster, K.M., Parkin, K.L., and Steele, J.L. (2000). Characterization of an arylesterase from *Lactobacillus helveticus* CNRZ32. *J. Appl. Microbiol.* 88**,** 572-583. doi: 10.1046/j.1365-2672.2000.00993.x.

Fenster, K.M., Parkin, K.L., and Steele, J.L. (2003b). Nucleotide sequencing, purification, and biochemical properties of an arylesterase from *Lactobacillus casei* LILA. *J. Dairy Sci.* 86**,** 2547-2557. doi: 10.3168/jds.S0022-0302(03)73849-1.

Fenster, K.M., Rankin, S.A., and Steele, J.L. (2003c). Accumulation of short n-chain ethyl esters by esterases of lactic acid bacteria under conditions simulating ripening Parmesan cheese. *J. Dairy Sci.* 86**,** 2818-2825. doi: 10.3168/jds.S0022-0302(03)73879-X.

Fernández De Palencia, P., De La Plaza, M., Amárita, F., Requena, T., and Peláez, C. (2006). Diversity of amino acid converting enzymes in wild lactic acid bacteria. *Enzyme Microb. Technol.* 38**,** 88-93. doi: 10.1016/j.enzmictec.2005.04.018.

Fernandez, M., and Zuniga, M. (2006). Amino acid catabolic pathways of lactic acid bacteria. *Crit. Rev. Microbiol.* 32**,** 155-183. doi: 10.1080/10408410600880643.

Fierer, N., Lauber, C.L., Zhou, N., Mcdonald, D., Costello, E.K., and Knight, R. (2010). Forensic identification using skin bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* 107**,** 6477. doi: 10.1073/pnas.1000162107.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269**,** 496-512. doi: 10.1126/science.7542800.

Folkenberg, D.M., Dejmek, P., Skriver, A., Skov Guldager, H., and Ipsen, R. (2006). Sensory and rheological screening of exopolysaccharide producing strains of bacterial yoghurt cultures. *Int. Dairy J.* 16**,** 111-118. doi: 10.1016/j.idairyj.2004.10.013.

Fonseca, N.A., Rung, J., Brazma, A., and Marioni, J.C. (2012). Tools for mapping high-throughput sequencing data. *Bioinformatics* 28**,** 3169-3177. doi: 10.1093/bioinformatics/bts605.

Food and Drug Administration [FDA] (2007). *21 CFR Part 131: Microorganisms & Microbial-Derived Ingredients Used in Food (Partial List)*. Silver Spring, MA:FDA.

Francois, Z., Félicité, M., and El Soda, M. (2004). Effect of ropy and capsular exopolysaccharides producing strain of *Lactobacillus plantarum* 162RM on characteristics and functionality of fermented milk and soft Kareish type cheese. *Afr. J. Biotechnol.* 3, 512-518. doi: 10.5897/AJB2004.000-2102.

Franz, C., Huch, M., Abriouel, H., and Gálvez, A. (2011). Enterococci as probiotics and their implications in food safety. *Int. J. Food Microbiol.* 151**,** 125-140. doi: 10.1016/j.ijfoodmicro.2011.08.014.

Franz, C.M.A.P., and Holzapfel, W.H. (2011). "The importance of understanding the stress physiology of lactic acid bacteria," in *Stress Responses of Lactic Acid Bacteria*, eds E. Tsakalidou, and K. Papadimitriou, (New York, NY: Springer Science+Business Media), 3–20. doi: 10.1007/978-0-387-92771-8_1

Freitas, M. (2017). "The benefits of yogurt, cultures, and fermentation," in *The Microbiota in Gastrointestinal Pathophysiology Implications for Human Health, Prebiotics, Probiotics, and Dysbiosis*, eds. M.H. Floch, Y. Ringel & W.A. Walker, (Boston: Academic Press), 209-223. doi: 10.1016/B978-0-12-804024-9.00024-0

Galia, W., Jameh, N., Perrin, C., Genay, M., and Dary-Mourot, A. (2016). Acquisition of PrtS in *Streptococcus thermophilus* is not enough in certain strains to achieve rapid milk acidification. *Dairy Sci. Technol.* 96, 623–636. doi: 10.1007/s13594-016-0292-3.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43**,** D261-269. doi: 10.1093/nar/gku1223.

Ganesan, B., Stuart, M.R., and Weimer, B.C. (2007). Carbohydrate starvation causes a metabolically active but nonculturable state in *Lactococcus lactis*. *Appl. Environ. Microbiol.* 73**,** 2498-2512. doi: 10.1128/aem.01832-06.

Gänzle, M.G., Vermeulen, N., and Vogel, R.F. (2007). Carbohydrate, peptide and lipid metabolism of lactic acid bacteria in sourdough. *Food Microbiol.* 24**,** 128-138. doi: 10.1016/j.fm.2006.07.006.

Gao, X.Y., Zhi, X.Y., Li, H.W., Klenk, H.P., and Li, W.J. (2014). Comparative genomics of the bacterial genus *Streptococcus* illuminates evolutionary implications of species groups. *PLoS One* 9**,** e101229. doi: 10.1371/journal.pone.0101229.

Garault, P., Letort, C., Juillard, V., and Monnet, V. (2000). Branched-chain amino acid biosynthesis is essential for optimal growth of *Streptococcus thermophilus* in milk. *Appl. Environ. Microbiol.* 66, 5128–5133. doi: 10.1128/aem.66.12.5128-5133.2000.

García Quintáns, N., Blancato, V., Repizo, G., Magni, C., and López, P. (2008). "Citrate metabolism and aroma compound production in lactic acid bacteria," in *Molecular Aspects of Lactic Acid Bacteria for Traditional and New Applications,* eds B. Mayo, P. López & G. Pérez-Martínez (Kerala: Research Signpost, India), 65-88.

Gardan, R., Besset, C., Guillot, A., Gitton, C., and Monnet, V. (2009). The oligopeptide transport system is essential for the development of natural competence in *Streptococcus thermophilus* strain LMD-9. *J. Bacteriol.* 191**,** 4647. doi: 10.1128/JB.00257-09.

Gatti, M., Bottari, B., Lazzi, C., Neviani, E., and Mucchetti, G. (2014). Invited review: Microbial evolution in raw-milk, long-ripened cheeses produced using undefined natural whey starters. *J. Dairy Sci.* 97**,** 573-591. doi: 10.3168/jds.2013-7187.

Gene Ontology Consortium (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43**,** D1049-D1056. doi: 10.1093/nar/gku1179.

George, F., Daniel, C., Thomas, M., Singer, E., Guilbaud, A., Tessier, F.J., et al. (2018). Occurrence and dynamism of lactic acid bacteria in distinct ecological niches: A multifaceted functional health perspective. *Front. Microbiol.* 9, 2899. doi: 10.3389/fmicb.2018.02899.

Germond, J.E., Lapierre, L., Delley, M., Mollet, B., Felis, G.E., and Dellaglio, F. (2003). Evolution of the bacterial species *Lactobacillus delbrueckii*: A partial genomic study with reflections on prokaryotic species concept. *Mol. Biol. Evol.* 20**,** 93-104. doi: 10.1093/molbev/msg012.

Giani, A.M., Gallo, G.R., Gianfranceschi, L., and Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput. Struct. Biotechnol. J.* 18**,** 9-19. doi: 10.1016/j.csbj.2019.11.002.

Giraffa, G., Andrighetto, C., Antonello, C., Gatti, M., Lazzi, C., Marcazzan, G., et al. (2004). Genotypic and phenotypic diversity of *Lactobacillus delbrueckii* subsp. *lactis* strains of dairy origin. *Int. J. Food Microbiol.* 91**,** 129-139. doi: 10.1016/s0168-1605(03)00368-4.

Glaasker, E., Heuberger, E.H., Konings, W.N., and Poolman, B. (1998). Mechanism of osmotic activation of the quaternary ammonium compound transporter (QacT) of *Lactobacillus plantarum*. *J. Bacteriol.* 180**,** 5540-5546. doi: 10.1128/jb.180.21.5540-5546.1998.

Gobbetti, M., De Angelis, M., Di Cagno, R., Mancini, L., and Fox, P. (2015). Pros and cons for using non-starter lactic acid bacteria (NSLAB) as secondary/adjunct starters for cheese ripening. *Trends Food Sci. Technol.* 45. doi: 10.1016/j.tifs.2015.07.016.

Gobbetti, M., Fox, P.F., and Stepaniak, L. (1996). Esterolytic and lipolytic activities of mesophilic and thermophilic lactobacilli. *Ital. J. Food Sci.* 2, 127-135.

Gobbetti, M., Fox, P.F., and Stepaniak, L. (1997a). Isolation and characterization of a tributyrin esterase from Lactobacillus plantaeum 2739. *J. Dairy Sci.* 80**,** 3099-3106. doi: 10.3168/jds.S0022-0302(97)76280-5.

Gobbetti, M., Smacchi, E., and Corsetti, A. (1997b). Purification and characterization of a cell surface-associated esterase from Lactoeacillus fermeetum DT41. *Int. Dairy J.* 7**,** 13-21. doi: 10.1016/S0958-6946(96)00025-8.

Goh, Y.J., Goin, C., O'flaherty, S., Altermann, E., and Hutkins, R. (2011). Specialized adaptation of a lactic acid bacterium to the milk environment: The comparative genomics of *Streptococcus thermophilus* LMD-9. *Microb. Cell Fact.* 10 (Suppl. 1):S22. doi: 10.1186/1475-2859-10-S1-S22.

Greenleaf, W.J., and Sidow, A. (2014). The future of sequencing: Convergence of intelligent design and market Darwinism. *Genome Biol.* 15**,** 303-303. doi: 10.1186/gb4168.

Griffiths, M.W., and Tellez, A.M. (2013). *Lactobacillus helveticus*: The proteolytic system. *Front. Microbiol.* 4**:**30. doi: 10.3389/fmicb.2013.00030.

Grobben, G.J., Van Casteren, W.H.M., Schols, H.A., Oosterveld, A., Sala, G., Smith, M.R., et al. (1997). Analysis of the exopolysaccharides produced by *Lactobacillus delbrueckii* subsp. *bulgaricus* NCFB 2772 grown in continuous culture on glucose and fructose. *Appl. Microbiol. Biotechnol.* 48**,** 516-521. doi: 10.1007/s002530051089.

Gruter, M., Leeflang, B.R., Kuiper, J., Kamerling, J.P., and Vliegenthart, J.F. (1993). Structural characterisation of the exopolysaccharide produced by *Lactobacillus delbrückii* subspecies

bulgaricus rr grown in skimmed milk. *Carbohydr. Res.* 239**,** 209-226. doi: 10.1016/0008-6215(93)84216-s.

Gu, S., Fang, L., and Xu, X. (2013). Using SOAPaligner for short reads alignment. *Curr. Protoc. Bioinformatics* 44. doi: 10.1002/0471250953.bi1111s44.

Guarner, F., Perdigon, G., Corthier, G., Salminen, S., Koletzko, B., and Morelli, L. (2007). Should yoghurt cultures be considered probiotic? *Br. J. Nutr.* 93**,** 783-786. doi: 10.1079/BJN20051428.

Hagting, A., Kunji, E.R., Leenhouts, K.J., Poolman, B., and Konings, W.N. (1994). The di- and tripeptide transport protein of *Lactococcus lactis*. A new type of bacterial peptide transporter. *J. Biol. Chem.* 269**,** 11391-11399.

Hammes, W.P, and Hertel, L. (2006). "The genera *Lactobacillus* and *Carnobacterium*," in *The Prokaryotes: Vol. 4: Bacteria: Firmicutes, Cyanobacteria*, eds M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt (New York: Springer), 320-403.

Hammes, W.P., and Vogel, R.F. (1995). "The genus *Lactobacillus*," in *The Genera of Lactic Acid Bacteria*, eds B.J.B. Wood & W.H. Holzapfel (London: Blackie Academic, UK), 19-54.

Hammes, W.P., Bantleon, A., and Min, S. (1990). Lactic acid bacteria in meat fermentation. *FEMS Microbiol. Lett.* 87**,** 165-173. doi: 10.1016/0378-1097(90)90704-T.

Hao, P., Zheng, H., Yu, Y., Ding, G., Gu, W., Chen, S., et al. (2011). Complete sequencing and pan-genomic analysis of *Lactobacillus delbrueckii* subsp. *bulgaricus* reveal its genetic basis for industrial yogurt production. *PLoS One* 6**,** e15964-e15964. doi: 10.1371/journal.pone.0015964.

Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., et al. (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* 10**,** R32. doi: 10.1186/gb-2009-10-3-r32.

Harris, T.D., Buzby, P.R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., et al. (2008). Single-molecule DNA sequencing of a viral genome. *Science* 320**,** 106-109. doi: 10.1126/science.1150427.

Heather, J.M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics* 107**,** 1-8. doi: 10.1016/j.ygeno.2015.11.003.

Hebert, E.M., Raya, R.R., Brown, L., Font De Valdez, G., Savoy De Giori, G., and Taranto, M.P. (2013). Genome sequence of the cheese-starter strain *Lactobacillus delbrueckii* subsp. *lactis* CRL 581. *Genome Announc.* 1**,** e00602-00613. doi: 10.1128/genomeA.00602-13.

Herve-Jimenez, L., Guillouard, I., Guedon, E., Boudebbouze, S., Hols, P., Monnet, V., et al. (2009). Postgenomic analysis of *Streptococcus thermophilus* cocultivated in milk with *Lactobacillus delbrueckii* subsp. *bulgaricus*: Involvement of nitrogen, purine, and iron metabolism. *Appl. Environ. Microbiol.* 75, 2062–2073. doi: 10.1128/AEM.01984-08

Hidalgo-Cantabrana, C., López, P., Gueimonde, M., De Los Reyes-Gavilán, C.G., Suárez, A., Margolles, A., et al. (2012). Immune modulation capability of exopolysaccharides synthesised by lactic acid bacteria and bifidobacteria. *Probiotics Antimicrob. Proteins* 4**,** 227-237. doi: 10.1007/s12602-012-9110-2.

Hols, P., Defrenne, C., Ferain, T., Derzelle, S., Delplace, B., and Delcour, J. (1997). The alanine racemase gene is essential for growth of *Lactobacillus plantarum*. *J. Bacteriol.* 179**,** 3804-3807. doi: 10.1128/jb.179.11.3804-3807.1997.

Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., et al. (2005). New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS Microbiol. Rev.* 29, 435–463. doi: 10.1016/j.fmrre.2005.04.008.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327**,** 167. doi: 10.1126/science.1179555.

Horvath, P., Coûté-Monvoisin, A.-C., Romero, D.A., Boyaval, P., Fremaux, C., and Barrangou, R. (2009). Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food Microbiol.* 131**,** 62-70. doi: 10.1016/j.ijfoodmicro.2008.05.030.

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., et al. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190**,** 1401. doi: 10.1128/JB.01415-07.

Hu, G., and Kurgan, L. (2019). Sequence similarity searching. *Curr. Protoc. Protein Sci.* 95**,** e71. doi: 10.1002/cpps.71.

Huang, S., He, J., Chang, S., Zhang, P., Liang, F., Li, S., et al. (2010). Identifying single bases in a DNA oligomer with electron tunnelling. *Nat. Nanotechnol.* 5**,** 868-873. doi: 10.1038/nnano.2010.213.

Huang, Y.F., Chen, S.C., Chiang, Y.S., Chen, T.H., and Chiu, K.P. (2012). Palindromic sequence impedes sequencing-by-ligation mechanism. *BMC Syst. Biol.* 6 (Suppl. 2):S10. doi: 10.1186/1752-0509-6-s2-s10.

Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11**,** 119. doi: 10.1186/1471-2105-11-119.

Irmler, S., Bavan, T., Oberli, A., Roetschi, A., Badertscher, R., Guggenbühl, B., et al. (2013). Catabolism of serine by *Pediococcus acidilactici* and *Pediococcus pentosaceus*. *Appl. Environ. Microbiol.* 79, 1309-1315. doi: 10.1128/AEM.03085-12.

Iskandar, C.F., Borges, F., Taminiau, B., Daube, G., Zagorec, M., Remenant, B., et al. (2017). Comparative genomic analysis reveals ecological differentiation in the genus *Carnobacterium*. *Front. Microbiol.* 8**,** 357. doi: 10.3389/fmicb.2017.00357.

Iskandar, C.F., Cailliez-Grimal, C., Borges, F., and Revol-Junelles, A.-M. (2019). Review of lactose and galactose metabolism in lactic acid bacteria dedicated to expert genomic annotation. *Trends Food Sci. Technol.* 88**,** 121-132. doi: 10.1016/j.tifs.2019.03.020.

Istace, B., Friedrich, A., D'agata, L., Faye, S., Payen, E., Beluche, O., et al. (2017). *De novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* 6**,** 1-13. doi: 10.1093/gigascience/giw018.

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36**,** 338-345. doi: 10.1038/nbt.4060.

Jolly, L., and Stingele, F. (2001). Molecular organization and functionality of exopolysaccharide gene clusters in lactic acid bacteria. *Int. Dairy J.* 11**,** 733-745. doi: 10.1016/S0958-6946(01)00117-0.

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'keefe, S., Mitchel, T., et al. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front. Microbiol.* 7**,** 459-459. doi: 10.3389/fmicb.2016.00459.

Kandler, O. (1983). Carbohydrate metabolism in lactic acid bacteria. *Antonie van Leeuwenhoek* 49**,** 209-224. doi: 10.1007/BF00399499.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28**,** 27-30. doi: 10.1093/nar/28.1.27.

Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29**,** 28-35. doi: 10.1002/pro.3711.

Kang, X., Ling, N., Sun, G., Zhou, Q., Zhang, L., and Sheng, Q. (2012). Complete genome sequence of *Streptococcus thermophilus* strain MN-ZLW-002. *J. Bacteriol.* 194, 4428–4429. doi: 10.1128/JB.00740-12.

Kanmani, P., Albarracin, L., Kobayashi, H., Hebert, E.M., Saavedra, L., Komatsu, R., et al. (2018). Genomic characterization of *Lactobacillus delbrueckii* TUA4408L and evaluation of the antiviral activities of its extracellular polysaccharides in porcine intestinal epithelial cells. *Front. Immunol.* 9. doi: 10.3389/fimmu.2018.02178.

Katz, M., Medina, R., Gonzalez, S., and Oliver, G. (2002). Esterolytic and lipolytic activities of lactic acid bacteria isolated from ewe's milk and cheese. *J. Food Prot.* 65**,** 1997-2001. doi: 10.4315/0362-028x-65.12.1997.

Kawamura, Y., Hou, X.G., Sultana, F., Miura, H., and Ezaki, T. (1995). Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *Int. J. Syst. Bacteriol.* 45**,** 406-408. doi: 10.1099/00207713-45-2-406.

Kieronczyk, A., Skeie, S., Langsrud, T., and Yvon, M. (2003). Cooperation between *Lactococcus lactis* and nonstarter lactobacilli in the formation of cheese aroma from amino acids. *Appl. Environ. Microbiol.* 69**,** 734-739. doi: 10.1128/aem.69.2.734-739.2003.

Kim, K.H., Lee, S.H., Chun, B.H., Jeong, S.E., and Jeon, C.O. (2019). *Tetragenococcus halophilus* MJ4 as a starter culture for repressing biogenic amine (cadaverine) formation during saeu-jeot (salted shrimp) fermentation. *Food Microbiol.* 82**,** 465-473. doi: 10.1016/j.fm.2019.02.017.

Kitazawa, H., Harata, T., Uemura, J., Saito, T., Kaneko, T., and Itoh, T. (1998). Phosphate group requirement for mitogenic activation of lymphocytes by an extracellular phosphopolysaccharide from *Lactobacillus delbrueckii* ssp. *bulgaricus*. *Int. J. Food Microbiol.* 40**,** 169-175. doi: 10.1016/s0168-1605(98)00030-0.

Kobayashi, J., Yukimoto, J., Shimizu, Y., Ohmori, T., Suzuki, H., Doi, K., et al. (2015). Characterization of *Lactobacillus salivarius* alanine racemase: Short-chain carboxylate-activation and the role of A131. *Springerplus* 4**,** 639-639. doi: 10.1186/s40064-015-1335-6.

Kongo, J.M. (2013). "Lactic acid bacteria as starter-cultures for cheese processing: past, present and future developments," in *Lactic Acid Bacteria - R & D for Food, Health and Livestock Purposes*, ed. J.M. Kongo, (London: IntechOpen), doi: 10.5772/55937.

Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., et al. (2012). Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30**,** 693-700. doi: 10.1038/nbt.2280.

Krzyściak, W., Pluskwa, K.K., Jurczak, A., and Kościelniak, D. (2013). The pathogenicity of the *Streptococcus* genus. *Eur. J. Clin. Microbiol. Infect. Dis.* 32**,** 1361-1376. doi: 10.1007/s10096-013-1914-9.

Kudo, Y., Oki, K., and Watanabe, K. (2012). *Lactobacillus delbrueckii* subsp. *sunkii* subsp. nov., isolated from sunki, a traditional Japanese pickle. *Int. J. Syst. Evol. Microbiol.* 62**,** 2643-2649. doi: 10.1099/ijs.0.037051-0.

Kwong, J.C., Mccallum, N., Sintchenko, V., and Howden, B.P. (2015). Whole genome sequencing in clinical and public health microbiology. *Pathology* 47**,** 199-210. doi: 10.1097/pat.0000000000000235.

Labrie, S.J., Tremblay, D.M., Plante, P.L., Wasserscheid, J., Dewar, K., Corbeil, J., et al. (2015). Complete genome sequence of *Streptococcus thermophilus* SMQ-301, a model strain for phage-host interactions. *Genome Announc.* 3. doi: 10.1128/genomeA.00480-15.

Lamont, J., Wilkins, O., Bywater-Ekegärd, M., and Smith, D. (2017). From yogurt to yield: Potential applications of lactic acid bacteria in plant production. *Soil Biol. Biochem.* 111**,** 1-9. doi: 10.1016/j.soilbio.2017.03.015.

Lamothe, G., Jolly, L., Mollet, B., and Stingele, F. (2002). Genetic and biochemical characterization of exopolysaccharide biosynthesis by *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Arch. Microbiol.* 178**,** 218-228. doi: 10.1007/s00203-002-0447-x.

Landete, J.M., Pardo, I., and Ferrer, S. (2006). Histamine, histidine, and growth-phase mediated regulation of the histidine decarboxylase gene in lactic acid bacteria isolated from wine. *FEMS Microbiol. Lett.* 260**,** 84-90. doi: 10.1111/j.1574-6968.2006.00294.x.

Lazzi, C., Povolo, M., Locci, F., Bernini, V., Neviani, E., and Gatti, M. (2016). Can the development and autolysis of lactic acid bacteria influence the cheese volatile fraction? The case of Grana Padano. *Int. J. Food Microbiol.* 233**,** 20-28. doi: 10.1016/j.ijfoodmicro.2016.06.009.

Le Bars, D., and Yvon, M. (2008). Formation of diacetyl and acetoin by *Lactococcus lactis* via aspartate catabolism. *J. Appl. Microbiol.* 104**,** 171-177. doi: 10.1111/j.1365-2672.2007.03539.x.

Lebeer, S., Vanderleyden, J., and De Keersmaecker, S.C.J. (2008). Genes and molecules of lactobacilli supporting probiotic action. *Microbiol. Mol. Biol. R.* 72**,** 728. doi: 10.1128/MMBR.00017-08.

Lebeer, S., Verhoeven, T.L.A., Francius, G., Schoofs, G., Lambrichts, I., Dufrêne, Y., et al. (2009). Identification of a gene cluster for the biosynthesis of a long, galactose-rich exopolysaccharide in *Lactobacillus rhamnosus* GG and functional analysis of the priming glycosyltransferase. *Appl. Environ. Microbiol.* 75**,** 3554-3563. doi: 10.1128/AEM.02919-08.

Leblanc, J.G., Aubry, C., Cortes-Perez, N.G., De Moreno De Leblanc, A., Vergnolle, N., Langella, P., et al. (2013). Mucosal targeting of therapeutic molecules using genetically modified lactic acid bacteria: An update. *FEMS Microbiol. Lett.* 344**,** 1-9. doi: 10.1111/1574-6968.12159.

Lee, N.-K., Hyun-Dong, P., Kyoung Jun, H., Si-Kyung, L., Su Jin, E., and Sung-Ho, S. (2015). Multifunctional effect of probiotic *Lactococcus lactis* KC24 isolated from kimchi. *LWT-Fodd Sci. Technol.* 64**,** 1036-1041. doi: 10.1016/j.lwt.2015.07.019.

Leech, J., Cabrera-Rubio, R., Walsh, A.M., Macori, G., Walsh, C.J., Barton, W., et al. (2020). Fermented food metagenomics reveals substrate-associated differences in taxonomy, health-associated- and antibiotic resistance-determinants. *bioRxiv***,** 2020.2003.2013.991653. doi: 10.1101/2020.03.13.991653.

Leroy, F., and De Vuyst, L. (2004). Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends Food Sci. Technol.* 15**,** 67-78. doi: 10.1016/j.tifs.2003.09.004.

Letort, C., and Juillard, V. (2001). Development of a minimal chemicallydefined medium for the exponential growth of *Streptococcus thermophilus*. *J. Appl. Microbiol.* 91, 1023–1029. doi: 10.1046/j.1365-2672.2001.01469.x.

Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., and Webb, W.W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299**,** 682-686. doi: 10.1126/science.1079700.

Li, B., Ding, X., Evivie, S.E., Jin, D., Meng, Y., Huo, G., et al. (2018). Short communication: Genomic and phenotypic analyses of exopolysaccharides produced by *Streptococcus thermophilus* KLDS SM. *J. Dairy Sci.* 101**,** 106-112. doi: 10.3168/jds.2017-13534.

Limauro, D., Falciatore, A., Basso, A. L., Forlani, G., and De Felice, M. (1996). Proline biosynthesis in *Streptococcus thermophilus*: Characterization of the proBA operon and its products. *Microbiology* 142, 3275–3282. doi: 10.1099/13500872-142-11-3275.

Linares, D.M., Arboleya, S., Ross, R.P., and Stanton, C. (2017). Complete genome sequence of the gamma-aminobutyric acid-producing strain *Streptococcus thermophilus* APC151. *Genome Announc.* 5**,** e00205-00217. doi: 10.1128/genomeA.00205-17.

Linares, D.M., O'callaghan, T.F., O'connor, P.M., Ross, R.P., and Stanton, C. (2016). *Streptococcus thermophilus* APC151 strain is suitable for the manufacture of naturally GABA-enriched bioactive yogurt. *Front. Microbiol.* 7**,** 1876-1876. doi: 10.3389/fmicb.2016.01876.

Liu, M., Bayjanov, J.R., Renckens, B., Nauta, A., and Siezen, R.J. (2010). The proteolytic system of lactic acid bacteria revisited: A genomic comparison. *BMC Genom.* 11**,** 36. doi: 10.1186/1471-2164-11-36.

Liu, M., Nauta, A., Francke, C., and Siezen, R.J. (2008). Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. *Appl. Environ. Microbiol.* 74**,** 4590-4600. doi: 10.1128/AEM.00150-08.

Liu, M., Siezen, R. J., and Nauta, A. (2009). *In silico* prediction of horizontal gene transfer events in *Lactobacillus bulgaricus* and *Streptococcus thermophilus* reveals protocooperation in yogurt manufacturing. *Appl. Environ.Microbiol.* 75,4120–4129. doi: 10.1128/AEM.02898-08

Liu, S.-Q., Holland, R., and Crow, V.L. (2001). Purification and properties of intracellular esterases from *Streptococcus thermophilus*. *Int. Dairy J.* 11**,** 27-35. doi: 10.1016/S0958-6946(01)00035-8.

Liu, S., Han, Y., and Zhou, Z.-J. (2011). Lactic acid bacteria in traditional fermented Chinese foods. *Food Res. Int.* 44**,** 643-651. doi: 10.1016/j.foodres.2010.12.034.

Liu, S.Q., Holland, R., and Crow, V.L. (2003a). The potential of dairy lactic acid bacteria to metabolise amino acids via non-transaminating reactions and endogenous transamination. *Int. J. Food Microbiol.* 86**,** 257-269. doi: 10.1016/s0168-1605(03)00040-0.

Liu, S.Q., Holland, R., Mcjarrow, P., and Crow, V.L. (2003b). Serine metabolism in *Lactobacillus plantarum*. *Int. J. Food Microbiol.* 89**,** 265-273. doi: 10.1016/S0168-1605(03)00157-0.

Liu, W., Yu, J., Sun, Z., Song, Y., Wang, X., Wang, H., et al. (2016). Relationships between functional genes in *Lactobacillus delbrueckii* ssp. *bulgaricus* isolates and phenotypic characteristics associated with fermentation time and flavor production in yogurt elucidated using multilocus sequence typing. *J. Dairy Sci.* 99**,** 89-103. doi: 10.3168/jds.2015-10209.

Loman, N.J., Constantinidou, C., Chan, J.Z., Halachev, M., Sergeant, M., Penn, C.W., et al. (2012a). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.* 10**,** 599-606. doi: 10.1038/nrmicro2850.

Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., et al. (2012b). Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 30**,** 434-439. doi: 10.1038/nbt.2198.

Loman, N.J., and Pallen, M.J. (2015). Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* 13**,** 787-794. doi: 10.1038/nrmicro3565.

Loman, N.J., Quick, J., and Simpson, J.T. (2015). A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat. Methods* 12, 733-735. doi: 10.1038/nmeth.3444.

Low, D., Ahlgren, J.A., Horne, D., Mcmahon, D.J., Oberg, C.J., and Broadbent, J.R. (1998). Role of *Streptococcus thermophilus* MR-1C capsular exopolysaccharide in cheese moisture retention. *Appl. Environ. Microbiol.* 64**,** 2147-2151.

Lucas, P.M., Blancato, V.S., Claisse, O., Magni, C., Lolkema, J.S., and Lonvaud-Funel, A. (2007). Agmatine deiminase pathway genes in *Lactobacillus brevis* are linked to the tyrosine decarboxylation operon in a putative acid resistance locus. *Microbiology* 153**,** 2221-2230. doi: 10.1099/mic.0.2007/006320-0.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1. doi: 10.1186/2047-217x-1-18.

Lynch, C.M., Mcsweeney, P.L.H., Fox, P., Cogan, T.M., and Drinan, F.D. (1997). Contribution of starter lactococci and non-starter lactobacilli to proteolysis in Cheddar cheese with a controlled microflora. *Lait* 77**,** 441-459.

Lynch, K.M., Mcsweeney, P.L.H., Arendt, E.K., Uniacke-Lowe, T., Galle, S., and Coffey, A. (2014). Isolation and characterisation of exopolysaccharide-producing *Weissella* and *Lactobacillus* and their application as adjunct cultures in Cheddar cheese. *Int. Dairy J.* 34**,** 125-134. doi: 10.1016/j.idairyj.2013.07.013.

Magni, C., De Mendoza, D., Konings, W.N., and Lolkema, J.S. (1999). Mechanism of citrate metabolism in *Lactococcus lactis*: Resistance against lactate toxicity at low pH. *J. Bacteriol.* 181**,** 1451-1457.

Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., et al. (2006). Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15611–15616. doi: 10.1073/pnas.0607117103.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., et al. (2020). Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18**,** 67-83. doi: 10.1038/s41579-019-0299-x.

Malang, S.K., Maina, N.H., Schwab, C., Tenkanen, M., and Lacroix, C. (2015). Characterization of exopolysaccharide and ropy capsular polysaccharide formation by *Weissella*. *Food Microbiol.* 46**,** 418-427. doi: 10.1016/j.fm.2014.08.022.

Manrao, E.A., Derrington, I.M., Laszlo, A.H., Langford, K.W., Hopper, M.K., Gillgren, N., et al. (2012). Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* 30**,** 349-353. doi: 10.1038/nbt.2171.

Mantegazza, C., Molinari, P., D'auria, E., Sonnino, M., Morelli, L., and Zuccotti, G.V. (2018). Probiotics and antibiotic-associated diarrhea in children: A review and new evidence on *Lactobacillus rhamnosus* GG during and after antibiotic treatment. *Pharmacol. Res.* 128**,** 63-72. doi: 10.1016/j.phrs.2017.08.001.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., Deweese-Scott, C., Geer, L.Y., Gwadz, M., et al. (2005). CDD: A Conserved Domain Database for protein classification. *Nucleic Acids Res.* 33**,** D192-196. doi: 10.1093/nar/gki069.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43**,** D222-226. doi: 10.1093/nar/gku1221.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437**,** 376-380. doi: 10.1038/nature03959.

Markowitz, V.M., Chen, I.M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: The Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* 40**,** D115-D122. doi: 10.1093/nar/gkr1044.

Martino, G.P., Quintana, I.M., Espariz, M., Blancato, V.S., and Magni, C. (2016). Aroma compounds generation in citrate metabolism of *Enterococcus faecium*: Genetic characterization of type I citrate gene cluster. *Int. J. Food Microbiol.* 218**,** 27-37. doi: 10.1016/j.ijfoodmicro.2015.11.004.

Matsuzaki, C., Kamishima, K., Matsumoto, K., Koga, H., Katayama, T., Yamamoto, K., et al. (2014). Immunomodulating activity of exopolysaccharide-producing *Leuconostoc mesenteroides* strain NTM048 from green peas. *J. Appl. Microbiol.* 116**,** 980-989. doi: 10.1111/jam.12411.

Mayo, B., Van Sinderen, D., and Ventura, M. (2008). Genome analysis of food grade lactic acid-producing bacteria: From basics to applications. *Curr. Genomics* 9, 169–183. doi: 10.2174/138920208784340731.

McCoy, R.C., Taylor, R.W., Blauwkamp, T.A., Kelley, J.L., Kertesz, M., Pushkarev, D., et al. (2014). Illumina TruSeq synthetic long-reads empower *de novo* assembly and resolve complex, highly-repetitive transposable elements. *PLoS One* 9**,** e106689. doi: 10.1371/journal.pone.0106689.

Mcleod, A., Mosleth, E.F., Rud, I., Branco Dos Santos, F., Snipen, L., Liland, K.H., et al. (2017). Effects of glucose availability in *Lactobacillus sakei*; metabolic change and regulation of the proteome and transcriptome. *PloS One* 12**,** e0187542-e0187542. doi: 10.1371/journal.pone.0187542.

McSweeney, P.L.H., Fox, P.F., and Ciocia, F. (2017). "Metabolism of residual lactose and of lactate and citrate," in *Cheese: Chemistry, Physics and Microbiology*, *Volume 1: General Aspects,* eds P. McSweeney, P.F. Fox, P. Cotter & D. Everett (London: Academic Press, UK), 411-421.

Médigue, C., and Moszer, I. (2007). Annotation, comparison and databases for hundreds of bacterial genomes. *Res. Microbiol.* 158**,** 724-736. doi: 10.1016/j.resmic.2007.09.009.

Michaylova, M., Minkova, S., Kimura, K., Sasaki, T., and Isawa, K. (2007). Isolation and characterization of *Lactobacillus delbrueckii* ssp. *bulgaricus* and *Streptococcus thermophilus* from plants in Bulgaria. *FEMS Microbiol. Lett.* 269**,** 160-169. doi: 10.1111/j.1574-6968.2007.00631.x.

Millen, A.M., Horvath, P., Boyaval, P., and Romero, D.A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in *Lactococcus lactis*. *PLoS One* 7**,** e51663. doi: 10.1371/journal.pone.0051663.

Mizuno, H., Tomotsune, K., Islam, M.A., Funabashi, R., Albarracin, L., Ikeda-Ohtsubo, W., et al. (2020). Exopolysaccharides from *Streptococcus thermophilus* ST538 modulate the antiviral innate immune response in porcine intestinal epitheliocytes. *Front. Microbiol.* 11**,** 894. doi: 10.3389/fmicb.2020.00894.

Mogna, L., Deidda, F., Nicola, S., Amoruso, A., Del Piano, M., and Mogna, G. (2016). *In vitro* inhibition of *Klebsiella pneumoniae* by *Lactobacillus delbrueckii* subsp. delbrueckii LDD01 (DSM 22106): An innovative strategy to possibly counteract such infections in humans? *J. Clin. Gastroenterol.* 50 (Suppl 2):S136-9. doi: 10.1097/mcg.0000000000000680.

Mokoena, M.P., Mutanda, T., and Olaniran, A.O. (2016). Perspectives on the probiotic potential of lactic acid bacteria from African traditional fermented foods and beverages. *Food Nutr. Res.* 60**,** 29630-29630. doi: 10.3402/fnr.v60.29630.

Monnet, C., Mora, D., and Corrieu, G. (2005). Glutamine synthesis is essential for growth of *Streptococcus thermophilus* in milk and is linked to urea catabolism. *Appl. Environ. Microbiol.* 71, 3376–3378. doi: 10.1128/AEM.71.6.3376-3378.2005.

Mora, D., Maguin, E., Masiero, M., Parini, C., Ricci, G., Manachini, P. L., et al. (2004). Characterization of urease genes cluster of *Streptococcus thermophilus*. *J. Appl. Microbiol.* 96, 209–219. doi: 10.1046/j.1365-2672.2003.02148.x.

Morovic, W., Hibberd, A.A., Zabel, B., Barrangou, R., and Stahl, B. (2016). Genotyping by PCR and high-throughput sequencing of commercial probiotic products reveals composition biases. *Front. Microbiol.* 7. doi: 10.3389/fmicb.2016.01747.

Mucchetti, G., Locci, F., Gatti, M., Neviani, E., Addeo, F., Dossena, A., et al. (2000). Pyroglutamic acid in cheese: Presence, origin, and correlation with ripening time of Grana Padano cheese. *J. Dairy Sci.* 83**,** 659-665. doi: 10.3168/jds.S0022-0302(00)74926-5.

Mucchetti, G., Locci, F., Massara, P., Vitale, R., and Neviani, E. (2002). Production of pyroglutamic acid by thermophilic lactic acid bacteria in hard-cooked mini-cheeses. *J. Dairy Sci.* 85**,** 2489-2496. doi: 10.3168/jds.S0022-0302(02)74331-2.

Müller, M., Lee, C.M., Gasiunas, G., Davis, T.H., Cradick, T.J., Siksnys, V., et al. (2016). *Streptococcus thermophilus* CRISPR-Cas9 systems enable specific editing of the human genome. *Mol. Ther.* 24**,** 636-644. doi: 10.1038/mt.2015.218.

Naidu, A.S., Bidlack, W.R., and Clemens, R.A. (1999). Probiotic spectra of lactic acid bacteria (LAB). *Crit. Rev. Food Sci. Nutr.* 39**,** 13-126. doi: 10.1080/10408699991279187.

Nannen, N.L., and Hutkins, R.W. (1991). Proton-translocating adenosine triphosphatase activity in lactic acid bacterial. *J. Dairy Sci.* 74**,** 747-751. doi: 10.3168/jds.S0022-0302(91)78220-9.

Neely, R.K., Deen, J., and Hofkens, J. (2011). Optical mapping of DNA: Single-molecule-based methods for mapping genomes. *Biopolymers* 95**,** 298-311. doi: 10.1002/bip.21579.

Nishimura, J., Kawai, Y., Aritomo, R., Ito, Y., Makino, S., Ikegami, S., et al. (2013). Effect of formic acid on exopolysaccharide production in skim milk fermentation by *Lactobacillus delbrueckii* subsp. *bulgaricus* OLL1073R-1. *Biosci. Microbiota Food Health* 32, 23–32. doi: 10.12938/bmfh.32.23.

Nout, M.J. (2009). Rich nutrition from the poorest-cereal fermentations in Africa and Asia. *Food Microbiol.* 26**,** 685-692. doi: 10.1016/j.fm.2009.07.002.

O'sullivan, D.J., Mcsweeney, P.L.H., Cotter, P.D., Giblin, L., and Sheehan, J.J. (2016). Compromised *Lactobacillus helveticus* starter activity in the presence of facultative heterofermentative *Lactobacillus casei* DPC6987 results in atypical eye formation in Swiss-type cheese. *J. Dairy Sci.* 99**,** 2625-2640. doi: 10.3168/jds.2015-10503.

O'sullivan, L., Ross, R.P., and Hill, C. (2002). Potential of bacteriocin-producing lactic acid bacteria for improvements in food safety and quality. *Biochimie* 84**,** 593-604. doi: 10.1016/s0300-9084(02)01457-8.

Olguin, N., Bordons, A., and Reguant, C. (2009). Influence of ethanol and pH on the gene expression of the citrate pathway in *Oenococcus oeni*. *Food Microbiol.* 26**,** 197-203. doi: 10.1016/j.fm.2008.09.004.

Onmus-Leone, F., Hang, J., Clifford, R.J., Yang, Y., Riley, M.C., Kuschner, R.A., et al. (2013). Enhanced *de novo* assembly of high throughput pyrosequencing data using whole genome mapping. *PLoS One* 8**,** e61762. doi: 10.1371/journal.pone.0061762.

Otero, M.C., Espeche, M.C., and Nader-Macías, M.E. (2007). Optimization of the freeze-drying media and survival throughout storage of freeze-dried *Lactobacillus gasseri* and *Lactobacillus delbrueckii* subsp. *delbrueckii* for veterinarian probiotic applications. *Process Biochem.* 42**,** 1406-1411. doi: 10.1016/j.procbio.2007.07.008.

Otero, M.C., Ocaña, V.S., and Elena Nader-Macías, M. (2004). Bacterial surface characteristics applied to selection of probiotic microorganisms. *Methods Mol. Biol.* 268**,** 435-440. doi: 10.1385/1-59259-766-1:435.

Ott, A., Germond, J., and Chaintreau, A. (2000). Origin of acetaldehyde during milk fermentation using 13 C-labeled precursors. *J. Agric. Food Chem.* 48**,** 1512-1517. doi: 10.1021/jf9904867.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., et al. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42**,** D206-D214. doi: 10.1093/nar/gkt1226.

Oyeniran, A., Gyawali,R., Aljaloud, S.O., Krastanov, A., Ibrahim, S.A. (2020). "Probiotic characteristics and health benefits of the yogurt bacterium *Lactobacillus delbrueckii*," in *Current Issues and Challenges in the Dairy Industry*, eds S.A. Ibrahim, T. Zimmerman & R. Gyawali, (London: IntechOpen), doi: 10.5772/intechopen.86939.

Padmanabhan, A., Tong, Y., Wu, Q., Zhang, J., and Shah, N.P. (2018). Transcriptomic insights into the growth phase- and sugar-associated changes in the exopolysaccharide production of a high EPS-producing *Streptococcus thermophilus* ASCC 1275. *Front. Microbiol.* 9**,** 1919. doi: 10.3389/fmicb.2018.01919.

Pagani, I., Liolios, K., Jansson, J., Chen, I.M.A., Smirnova, T., Nosrat, B., et al. (2012). The Genomes OnLine Database (GOLD) v.4: Status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* 40**,** D571-D579. doi: 10.1093/nar/gkr1100.

Pallen, M.J. (2016). Microbial bioinformatics 2020. *Microb. Biotechnol.* 9**,** 681-686. doi: 10.1111/1751-7915.12389.

Palmer, K.L., and Gilmore, M.S. (2010). Multidrug-resistant enterococci lack CRISPR-cas. *mBio* 1. doi: 10.1128/mBio.00227-10.

Palomba, S., Cavella, S., Torrieri, E., Piccolo, A., Mazzei, P., Blaiotta, G., et al. (2012). Polyphasic screening, homopolysaccharide composition, and viscoelastic behavior of wheat sourdough from a *Leuconostoc lactis* and *Lactobacillus curvatus* exopolysaccharide-producing starter culture. *Appl. Environ. Microbiol.* 78**,** 2737-2747. doi: 10.1128/aem.07302-11.

Palumbo, E., Favier, C.F., Deghorain, M., Cocconcelli, P.S., Grangette, C., Mercenier, A., et al. (2004). Knockout of the alanine racemase gene in *Lactobacillus plantarum* results in septation defects and cell wall perforation. *FEMS Microbiol. Lett.* 233**,** 131-138. doi: 10.1016/j.femsle.2004.02.001.

Pang, X., Li, W., Yang, L., Hu, C., Lu, J., Zhang, S., et al. (2019). Whole-genome sequencing and genomic-based acid tolerance mechanisms of *Lactobacillus delbrueckii* subsp. *bulgaricus LJJ. Preprint (Version 1) available at Research Square*. doi: 10.21203/rs.2.16590/v1.

Papadimitriou, K., Alegría, Á., Bron, P.A., De Angelis, M., Gobbetti, M., Kleerebezem, M., et al. (2016). Stress physiology of lactic acid bacteria. *Microbiol. Mol. Biol. Rev.* 80 (3)**,** 837-890. doi: 10.1128/MMBR.00076-15.

Papadimitriou, K., Alexandraki, V., and Tsakalidou, E. (2014). "Fighting off human infections," in *Interactive Probiotics*, ed. E. Pessione, (Boca Raton, FL: CRC Press), 22-51.

Papadimitriou, K., Pot, B., and Tsakalidou, E. (2015). How microbes adapt to a diversity of food niches. *Curr. Opin. Food Sci.* 2, 29–35. doi: 10.1016/j.cofs.2015.01.001.

Parente, E., Cogan, T.M., and Powell, I.B. (2017). "Starter cultures: General aspects," in *Cheese: Chemistry, Physics and Microbiology*, *Volume 1: General Aspects,* eds P. McSweeney, P.F. Fox, P. Cotter & D. Everett (London: Academic Press, UK), 201-226.

Pastink, M. I., Teusink, B., Hols, P., Visser, S., De Vos, W. M., and Hugenholtz, J. (2009). Genome-scale model of *Streptococcus thermophilus* LMG18311 for metabolic comparison of lactic acid bacteria. *Appl. Environ. Microbiol.* 75, 3627–3633. doi: 10.1128/AEM.00138-09.

Pati, A., Ivanova, N.N., Mikhailova, N., Ovchinnikova, G., Hooper, S.D., Lykidis, A., and Kyrpides, N.C. (2010). GenePRIMP: A gene prediction improvement pipeline for prokaryotic genomes. *Nat. Methods* 7**,** 455-457. doi: 10.1038/nmeth.1457.

Payne, A., Holmes, N., Rakyan, V., and Loose, M. (2019). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35**,** 2193-2198. doi: 10.1093/bioinformatics/bty841.

Péant, B., LaPointe, G., Gilbert, C., Atlan, D., Ward, P., and Roy, D. (2005). Comparative analysis of the exopolysaccharide biosynthesis gene clusters from four strains of *Lactobacillus rhamnosus*. *Microbiology* 151**,** 1839-1851. doi: 10.1099/mic.0.27852-0.

Pedersen, M.B., Gaudu, P., Lechardeur, D., Petit, M.A., and Gruss, A. (2012). Aerobic respiration metabolism in lactic acid bacteria and uses in biotechnology. *Annu. Rev. Food Sci. Technol.* 3**,** 37-58. doi: 10.1146/annurev-food-022811-101255.

Peltoniemi, K., Vesanto, E., and Palva, A. (2002). Genetic characterization of an oligopeptide transport system from *Lactobacillus* delbrueckii subsp. *bulgaricus*. *Arch. Microbiol.* 177**,** 457-467. doi: 10.1007/s00203-002-0411-9.

Peralta, G.H., Bergamini, C.V., and Hynes, E.R. (2016). Aminotransferase and glutamate dehydrogenase activities in lactobacilli and streptococci. *Braz. J. Microbiol.* 47**,** 741-748. doi: 10.1016/j.bjm.2016.04.005.

Peralta, G.H., Bergamini, C.V., Wolf, I.V., Perotti M.C., and Hynes E.R. (2018), "Adjunct cultures from non-starter lactic acid bacteria," in Cheese-Production, Consumption and Health Benefits, eds M.H.F. Henriques & C.J.D. Pereira, (New York, NY: Nova Science Publishers, Inc.), 39-96.

Pessione, E., and Cirrincione, S. (2016). Bioactive molecules released in food by lactic acid bacteria: Encrypted peptides and biogenic amines. *Front. Microbiol.* 7**,** 876. doi: 10.3389/fmicb.2016.00876.

Petry, S., Furlan, S., Waghorne, E., Saulnier, L., Cerning Dagger, J., and Maguin, E. (2003). Comparison of the thickening properties of four *Lactobacillus delbrueckii* subsp. *bulgaricus* strains and physicochemical characterization of their exopolysaccharides. *FEMS Microbiol. Lett.* 221**,** 285-291. doi: 10.1016/s0378-1097(03)00214-3.

Pightling, A.W., Petronella, N., and Pagotto, F. (2014). Choice of reference sequence and assembler for alignment of *Listeria monocytogenes* short-read sequence data greatly influences rates of error in SNP analyses. *PLoS One* 9**,** e104579. doi: 10.1371/journal.pone.0104579.

Pokusaeva, K., Fitzgerald, G.F., and Van Sinderen, D. (2011). Carbohydrate metabolism in Bifidobacteria. *Genes Nutr.* 6**,** 285-306. doi: 10.1007/s12263-010-0206-6.

Poptsova, M.S., and Gogarten, J.P. (2010). Using comparative genome analysis to identify problems in annotated microbial genomes. *Microbiology* 156**,** 1909-1917. doi: 10.1099/mic.0.033811-0.

Poyart, C., Quesne, G., Coulon, S., Berche, P., and Trieu-Cuot, P. (1998). Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *J. Clin. Microbiol.* 36**,** 41-47.

Prasad, J., Mcjarrow, P., and Gopal, P. (2003). Heat and osmotic stress responses of probiotic *Lactobacillus rhamnosus* HN001 (DR20) in relation to viability after drying. *Appl. Environ. Microbiol.* 69**,** 917-925. doi: 10.1128/aem.69.2.917-925.2003.

Punina, N.V., Makridakis, N.M., Remnev, M.A., and Topunov, A.F. (2015). Whole-genome sequencing targets drug-resistant bacterial infections. *Hum. Genomics* 9**,** 19-19. doi: 10.1186/s40246-015-0037-z.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., et al. (2012). A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genom.* 13**,** 341. doi: 10.1186/1471-2164-13-341.

Rahbar Saadat, Y., Yari Khosroushahi, A., and Pourghassem Gargari, B. (2019). A comprehensive review of anticancer, immunomodulatory and health beneficial effects of the lactic acid bacteria exopolysaccharides. *Carbohydr. Polym.* 217**,** 79-89. doi: 10.1016/j.carbpol.2019.04.025.

Ramraj, S.K., Paul, P., Neelakandan, Y., Paari, K., Pattukumar, V., and Arul, V. (2012). Traditional Indian fermented foods: A rich source of lactic acid bacteria. *Int. J. Food Sci. Nutr.* 64. doi: 10.3109/09637486.2012.746288.

Rantsiou, K., Urso, R., Dolci, P., Comi, G., and Cocolin, L. (2008). Microflora of Feta cheese from four Greek manufacturers. *Int. J. Food Microbiol.* 126**,** 36-42. doi: 10.1016/j.ijfoodmicro.2008.04.031.

Rasko, D.A., Worsham, P.L., Abshire, T.G., Stanley, S.T., Bannan, J.D., Wilson, M.R., et al. (2011). *Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. *Proc. Natl. Acad. Sci. U.S.A.* 108**,** 5027. doi: 10.1073/pnas.1016657108.

Renye, J.A., Jr., Needleman, D.S., Somkuti, G.A., and Steinberg, D.H. (2017). Complete genome sequence of *Streptococcus thermophilus* Strain B59671, which naturally produces the broad-spectrum bacteriocin thermophilin 110. *Genome Announc.* 5. doi: 10.1128/genomeA.01213-17.

Ricci, A., Allende, A., Bolton, D., Chemaly, M., Davies, R., Girones, R., et al. (2017). Scientific opinion on the update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to *EFSA. EFSA J.* 15:4664.

Richards, V.P., Palmer, S.R., Pavinski Bitar, P.D., Qin, X., Weinstock, G.M., Highlander, S.K., et al (2014). Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. *Genome Biol. Evol.* 6**,** 741-753. doi: 10.1093/gbe/evu048.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., et al. (2010). *De novo* assembly and analysis of RNA-seq data. *Nat. Methods* 7**,** 909-912. doi: 10.1038/nmeth.1517.

Rolny, I.S., Tiscornia, I., Racedo, S.M., Pérez, P.F., and Bollati-Fogolín, M. (2016). *Lactobacillus delbrueckii* subsp *lactis* CIDCA 133 modulates response of human epithelial and dendritic cells infected with *Bacillus cereus*. *Benef. Microbes* 7**,** 749-760. doi: 10.3920/bm2015.0191.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242**,** 84-89. doi: 10.1006/abio.1996.0432.

Ronaghi, M., Uhlén, M., and Nyrén, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* 281**,** 363, 365. doi: 10.1126/science.281.5375.363.

Ronholm, J., Nasheri, N., Petronella, N., and Pagotto, F. (2016). Navigating microbiological food safety in the era of whole-genome sequencing. *Clin. Microbiol. Rev.* 29**,** 837-857. doi: 10.1128/cmr.00056-16.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475**,** 348-352. doi: 10.1038/nature10242.

Ruas-Madiedo, P., Hugenholtz, J., and Zoon, P. (2002). An overview of the functionality of exopolysaccharides produced by lactic acid bacteria. *Int. Dairy J.* 12**,** 163-171. doi: https://doi.org/10.1016/S0958-6946(01)00160-1.

Russo, P., De La Luz Mohedano, M., Capozzi, V., De Palencia, P.F., López, P., Spano, G., et al. (2012). Comparative proteomic analysis of *Lactobacillus plantarum* WCFS1 and ΔctsR mutant strains under physiological and heat stress conditions. *Int. J. Mol. Sci.* 13, 10680-10696. doi: 10.3390/ijms130910680.

Ryan, P., Ross, P., Fitzgerald, G., Caplice, N., and Stanton, C. (2015). Sugar-coated: Exopolysaccharide producing lactic acid bacteria for food and human health applications. *Food Funct.* 6, 679-693. doi: 10.1039/C4FO00529E.

Rychlik, M., and Bosset, J. (2001). Flavour and off-flavour compounds of Swiss Gruyère cheese. Evaluation of potent odorants. *Int. Dairy J.* 11**,** 895-901. doi: 10.1016/S0958-6946(01)00108-X.

Salvetti, E., Harris, H.M.B., Felis, G.E., and Toole, P.W. (2018). Comparative genomics of the genus *Lactobacillus* reveals robust phylogroups that provide the basis for reclassification. *Appl. Environ. Microbiol.* 84**,** e00993-00918. doi: 10.1128/AEM.00993-18.

Salvetti, E., and O'Toole, P.W. (2017). When regulation challenges innovation: The case of the genus *Lactobacillus*. *Trends Food Sci. Technol.* 66**,** 187-194. doi: 10.1016/j.tifs.2017.05.009.

Sampson, T.R., and Weiss, D.S. (2014). CRISPR-Cas systems: New players in gene regulation and bacterial physiology. *Front. Cell Infect. Microbiol.* 4**,** 37. doi: 10.3389/fcimb.2014.00037.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74**,** 5463-5467. doi: 10.1073/pnas.74.12.5463.

Sanz, Y., Lanfermeijer, F.C., Renault, P., Bolotin, A., Konings, W.N., and Poolman, B. (2001). Genetic and functional characterization of dpp genes encoding a dipeptide transport system in *Lactococcus lactis*. *Arch. Microbiol.* 175**,** 334-343. doi: 10.1007/s002030100270.

Sarantinopoulos, P., Kalantzopoulos, G., and Tsakalidou, E. (2001). Citrate metabolism by *Enterococcus faecalis* FAIR-E 229. *Appl. Environ. Microbiol.* 67**,** 5482-5487. doi: 10.1128/AEM.67.12.5482-5487.2001.

Sasaki, Y., Horiuchi, H., *Kawashima*, H., Mukai, T., and Yamamoto, Y. (2014). NADH oxidase of *Streptococcus* thermophilus 1131 is required for the effective yogurt fermentation with *Lactobacillus delbrueckii* subsp. *bulgaricus* 2038. *Biosci. Microbiota Food Health* 33, 31-40. doi: 10.12938/bmfh.33.31.

Sauvageot, N., Beaufils, S., Mazé, A., Deutscher, J., and Hartke, A. (2006). Cloning and characterization of a gene encoding a cold-shock protein in *Lactobacillus casei*. *FEMS Microbiol. Lett.* 254**,** 55-62. doi: 10.1111/j.1574-6968.2005.00006.x.

Savijoki, K., Ingmer, H., and Varmanen, P. (2006). Proteolytic systems of lactic acid bacteria. *Appl. Microbiol. Biotechnol.* 71**,** 394-406. doi: 10.1007/s00253-006-0427-1.

Senne, M.M., and Gilliland, S.E. (2003). Antagonistic action of cells of *Lactobacillus delbrueckii* subsp. *lactis* against pathogenic and spoilage microorganisms in fresh meat systems. *J. Food Prot.* 66**,** 418-425. doi: 10.4315/0362-028x-66.3.418.

Settachaimongkon, S., Nout, M. J., Antunes Fernandes, E. C., Hettinga, K. A., Vervoort, J. M., Van Hooijdonk, T. C., et al. (2014). Influence of different proteolytic strains of *Streptococcus thermophilus* in co-culture with *Lactobacillus delbrueckii* subsp. *bulgaricus* on the metabolite profile of set-yoghurt. *Int. J. Food Microbiol.* 177, 29–36. doi: 10.1016/j.ijfoodmicro.2014.02.008.

Settanni, L., and Moschetti, G. (2010). Non-starter lactic acid bacteria used to improve cheese quality and provide health benefits. *Food Microbiol.* 27**,** 691-697. doi: 10.1016/j.fm.2010.05.023.

Sgarbi, E., Lazzi, C., Tabanelli, G., Gatti, M., Neviani, E., and Gardini, F. (2013). Nonstarter lactic acid bacteria volatilomes produced using cheese components. *J. Dairy Sci.* 96**,** 4223-4234. doi: 10.3168/jds.2012-6472.

Shi, Y., Chen, Y., Li, Z., Yang, L., Chen, W., and Mu, Z. (2015). Complete genome sequence of *Streptococcus thermophilus* MN-BM-A02, a rare strain with a high acid-producing rate and low post-acidification ability. *Genome Announc.* 3. doi: 10.1128/genomeA.00979-15.

Sieuwerts, S., Molenaar, D., Van Hijum, S. A., Beerthuyzen, M., Stevens, M. J., Janssen, P. W., et al. (2010). Mixed-culture transcriptome analysis reveals the molecular basis of mixed-culture growth in *Streptococcus thermophilus* and *Lactobacillus bulgaricus*. *Appl. Environ. Microbiol.* 76, 7775–7784. doi: 10.1128/AEM.01122-10.

Sigrist, C.J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., et al. (2002). PROSITE: A documented database using patterns and profiles as motif descriptors. *Brief Bioinform.* 3**,** 265-274. doi: 10.1093/bib/3.3.265.

Silvi, S., Nardi, M., Sulpizio, R., Orpianesi, C., Caggiano, M., Carnevali, O., et al. (2008). Effect of the addition of *Lactobacillus delbrueckii* subsp. *delbrueckii* on the gut microbiota composition and contribution to the well-being of European sea bass (Dicentrarchus labrax, L.). *Microb. Ecol. Health Dis.* 20**,** 53-59. doi: 10.1080/08910600701876020.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19**,** 1117-1123. doi: 10.1101/gr.089532.108.

Singh, T.K., Drake, M.A., and Cadwallader, K.R. (2003). Flavor of Cheddar cheese: A chemical and sensory perspective. *Compr. Rev. Food Sci. Food Saf.* 2**,** 166-189. doi: 10.1111/j.1541-4337.2003.tb00021.x.

Skeie, S., Kieronczyk, A., Næss, R.M., and Østlie, H. (2008). *Lactobacillus* adjuncts in cheese: Their influence on the degradation of citrate and serine during ripening of a washed curd cheese. *Int. Dairy J.* 18**,** 158-168. doi: 10.1016/j.idairyj.2007.09.003.

Slattery, L., O'callaghan, J., Fitzgerald, G.F., Beresford, T., and Ross, R.P. (2010). Invited review: *Lactobacillus helveticus*-a thermophilic dairy starter related to gut bacteria. *J. Dairy Sci.* 93**,** 4435-4454. doi: 10.3168/jds.2010-3327.

Song, Y., Sun, Z., Guo, C., Wu, Y., Liu, W., Yu, J., et al. (2016). Genetic diversity and population structure of *Lactobacillus delbrueckii* subspecies *bulgaricus* isolated from naturally fermented dairy foods. *Sci. Rep.* 6**,** 22704-22704. doi: 10.1038/srep22704.

Spano, G., Russo, P., Lonvaud-Funel, A., Lucas, P., Alexandre, H., Grandvalet, C., et al. (2010). Biogenic amines in fermented foods. *Eur. J. Clin. Nutr.* 64, S95-100. doi: 10.1038/ejcn.2010.218.

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. *Nucleic Acids Res.* 6**,** 2601-2610. doi: 10.1093/nar/6.7.2601.

Steen, A., Palumbo, E., Deghorain, M., Cocconcelli, P.S., Delcour, J., Kuipers, O.P., et al. (2005). Autolysis of *Lactococcus lactis* is increased upon D-alanine depletion of peptidoglycan and lipoteichoic acids. *J. Bacteriol.* 187**,** 114-124. doi: 10.1128/jb.187.1.114-124.2005.

Sumby, K.M., Matthews, A.H., Grbin, P.R., and Jiranek, V. (2009). Cloning and characterization of an intracellular esterase from the wine-associated lactic acid bacterium *Oenococcus oeni*. *Appl. Environ. Microbiol.* 75**,** 6729-6735. doi: 10.1128/AEM.01563-09.

Sun, Z., Chen, X., Wang, J., Zhao, W., Shao, Y., Guo, Z., et al. (2011a). Complete genome sequence of Lactobacillus delbrueckii subsp. bulgaricus strain ND02. *J. Bacteriol.* 193**,** 3426-3427. doi: 10.1128/JB.05004-11.

Sun, Z., Chen, X., Wang, J., Zhao, W., Shao, Y., Wu, L., et al. (2011b). Complete genome sequence of *Streptococcus thermophilus* strain ND03. *J. Bacteriol.* 193**,** 793-794. doi: 10.1128/JB.01374-10.

Sun, Z., Harris, H.M.B., Mccann, A., Guo, C., Argimón, S., Zhang, W., et al. (2015). Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat. Commun.* 6**,** 8322. doi: 10.1038/ncomms9322.

Suokko, A., Poutanen, M., Savijoki, K., Kalkkinen, N., and Varmanen, P. (2008). ClpL is essential for induction of thermotolerance and is potentially part of the HrcA regulon in *Lactobacillus gasseri*. *Proteomics* 8, 1029-1041. doi: 10.1002/pmic.200700925.

Šušković, J., Kos, B., Novak, J., Pavunc, A., Habjanič, K., and Matoć, S. (2010). Antimicrobial activity - The most important property of probiotic and starter lactic acid bacteria. *Food Technol. Biotechnol.* 48**,** 296-307.

Suzuki, C., Kobayashi, M., and Kimoto-Nira, H. (2013). Novel exopolysaccharides produced by *Lactococcus lactis* subsp. *lactis*, and the diversity of epsE genes in the exopolysaccharide biosynthesis gene clusters. *Biosci. Biotechnol. Biochem.* 77**,** 2013-2018. doi: 10.1271/bbb.130322.

Tagg, J.R., Wescombe, P.A., Hale, J.D.F., and Burton, J.P. (2019). "*Streptococcus*: a brief update on the current taxonomic status of the genus," in *Lactic Acid Bacteria Microbiological and Functional Aspects*, eds G. Vinderola, A. Ouwehand, S. Salminen & A. Wright (Boca Raton, FL: CRC Press), 87-108.

Tammam, J.D., Williams, A.G., Noble, J., and Lloyd, D. (2000). Amino acid fermentation in non-starter *Lactobacillus* spp. isolated from cheddar cheese. *Lett. Appl. Microbiol.* 30**,** 370-374. doi: 10.1046/j.1472-765x.2000.00727.x.

Tanasupawat, S., Phoottosavako, M., and Keeratipibul, S. (2015). Characterization and lipolytic activity of lactic acid bacteria isolated from Thai fermented meat. *J. Appl. Pharm. Sci.* 5**,** 006-012. doi: 10.7324/JAPS.2015.50302.

Täpp, J., Thollesson, M., and Herrmann, B. (2003). Phylogenetic relationships and genotyping of the genus *Streptococcus* by sequence determination of the RNase P RNA gene, rnpB. *Int. J. Syst. Evol. Microbiol.* 53**,** 1861-1871. doi: 10.1099/ijs.0.02639-0.

Tarrah, A., Castilhos, J.D., Rossi, R.C., Duarte, V.D.S., Ziegler, D.R., Corich, V., et al. (2018). *In vitro* probiotic potential and anti-cancer activity of newly isolated folate-producing *Streptococcus thermophilus* Strains. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.02214.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997). A genomic perspective on protein families. *Science* 278**,** 631-637. doi: 10.1126/science.278.5338.631.

Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44**,** 6614-6624. doi: 10.1093/nar/gkw569.

Tharrington, G., and Sorrells, K.M. (1992). Inhibition of *Listeria monocytogenes* by milk culture filtrates from *Lactobacillus delbrueckii* subsp. *lactis*. *J. Food Prot.* 55**,** 542-544. doi: 10.4315/0362-028x-55.7.542.

Thierry, A., Pogačić, T., Weber, M., and Lortal, S. (2015). "Production of flavor compounds by lactic acid bacteria in fermented foods," in *Biotechnology of Lactic Acid Bacteria: Novel Applications*, eds F. Mozzi, R.R. Raya & G.M. Vignolo (West Sussex: Wiley-Blackwell, UK), 314-340.

Thompson, A., Griffin, H., and Gasson, M.J. (2002). Characterization of an alanine racemase gene from *Lactobacillus reuteri*. *Curr. Microbiol.* 44**,** 246-250. doi: 10.1007/s00284-001-0030-8.

Thompson, J. (1987). Regulation of sugar transport and metabolism in lactic acid bacteria. *FEMS Microbiol. Lett.* 46**,** 221-231. doi: 10.1016/0378-1097(87)90109-1.

Thongsanit, J., Tanikawa, M., Yano, S., Tpachiki, T., and Wakayam, M. (2009). Identification of glutaminase-producing lactic acid bacteria isolated from Nham, a traditional Thai fermented food and characterisation of glutaminase activity of isolated *Weissella cibaria*. *Ann. Microbiol.* 59**,** 715-720. doi: 10.1007/BF03179213.

Tian, H., Li, B., Evivie, S. E., Sarker, S. K., Chowdhury, S., Lu, J., et al. (2018). Technological and genomic analysis of roles of the cell-envelope protease PrtS in yoghurt starter development. *Int. J. Mol. Sci.* 19. doi: 10.3390/ijms19041068.

Tomita, H., and Ike, Y. (2005). Genetic analysis of transfer-related regions of the vancomycin resistance *Enterococcus* conjugative plasmid pHTbeta: Identification of oriT and a putative relaxase gene. *J. Bacteriol.* 187**,** 7727-7737. doi: 10.1128/jb.187.22.7727-7737.2005.

Tsakalidou, E., Dalezios, I., and Kalantzopoulos, G. (1994a). Isolation and partial characterization of an intracellular esterase from *Enterococcus faecium* ACA-DC 237. *J. Biotechnol.* 37**,** 201-208. doi: 10.1016/0168-1656(94)90127-9.

Tsakalidou, E., and Kalantzopoulos, G. (1992). Purification and partial characterization of an esterase from *Lactococcus lactis* ssp *lactis* strain ACA-DC 127. *Le Lait* 72**,** 533-543. doi: 10.1051/lait:1992638.

Tsakalidou, E., Manolopoulou, E., Kabaraki, E., Zoidou, E., Pot, B., Kersters, K., et al. (1994b). The combined use of whole-cell protein extracts for the identification (SDS-PAGE) and enzyme activity screening of lactic acid bacteria isolated from rraditional Greek dairy products. *Syst. Appl. Microbiol.* 17**,** 444-458. doi: 10.1016/S0723-2020(11)80062-7.

Tufail, M., Hussain, S., Malik, F., Mirza, T., Parveen, G., Shafaat, S., et al. (2011). Isolation and evaluation of antibacterial activity of bacteriocin produced by *Lactobacillus bulgaricus* from yogurt. *Afr. J. Microbiol. Res.* 5**,** 3842-3847.

Tynkkynen, S., Buist, G., Kunji, E., Kok, J., Poolman, B., Venema, G., and Haandrikman, A. (1993). Genetic and biochemical characterization of the oligopeptide transport system of *Lactococcus lactis*. *J. Bacteriol.* 175**,** 7523-7532. doi: 10.1128/jb.175.23.7523-7532.1993.

Uchiyama, I. (2007). MBGD: A platform for microbial comparative genomics based on the automated construction of orthologous groups. *Nucleic Acids Res.* 35**,** D343-D346. doi: 10.1093/nar/gkl978.

Urbach, G. (1995). Contribution of lactic acid bacteria to flavour compound formation in dairy products. *Int. Dairy J.* 5**,** 877-903. doi: https://doi.org/10.1016/0958-6946(95)00037-2.

Uriot, O., Denis, S., Junjua, M., Roussel, Y., Dary-Mourot, A., and Blanquet-Diot, S. (2017). *Streptococcus thermophilus*: From yogurt starter to a new promising probiotic candidate? *J. Funct. Foods* 37**,** 74-89. doi: https://doi.org/10.1016/j.jff.2017.07.038.

Utturkar, S.M., Klingeman, D.M., Land, M.L., Schadt, C.W., Doktycz, M.J., Pelletier, D.A., et al. (2014). Evaluation and validation of *de novo* and hybrid assembly techniques to derive high-quality genome sequences. *Bioinformatics* 30**,** 2709-2716. doi: 10.1093/bioinformatics/btu391.

Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., et al. (2020). MicroScope: An integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res.* 48**,** D579-d589. doi: 10.1093/nar/gkz926.

Vallenet, D., Engelen, S., Mornico, D., Cruveiller, S., Fleury, L., Lajus, A., et al. (2009). MicroScope: A platform for microbial genome annotation and comparative genomics. *Database* 2009. doi: 10.1093/database/bap021.

van De Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., et al. (2006). The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc. Natl. Acad. Sci. U.S.A.* 103**,** 9274-9279. doi: 10.1073/pnas.0603024103.

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30**,** 418-426. doi: 10.1016/j.tig.2014.07.001.

van Kranenburg, R., Kleerebezem, M., van Hylckama Vlieg, J., Ursing, B.M., Boekhorst, J., Smit, B.A., et al. (2002). Flavour formation from amino acids by lactic acid bacteria: Predictions from genome sequence analysis. *Int. Dairy J.* 12**,** 111-121. doi: 10.1016/S0958-6946(01)00132-7.

van Kranenburg, R.V., Marugg, J.D., Van Swam, I.I., Willem, N.J., and De Vos, W.M. (1997). Molecular characterization of the plasmid-encoded eps gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis*. *Mol. Microbiol.* 24**,** 387-397. doi: 10.1046/j.1365-2958.1997.3521720.x.

von Wright, A., and Axelsson, L. (2019). "Lactic acid bacteria: An Introduction," in *Lactic Acid Bacteria Microbiolgical and Functional Aspects* eds G. Vinderola, A. Ouwehand, S. Salminen & A. Wright (Boca Raton, FL: CRC Press), 1-16.

Vaughan, E.E., Van Den Bogaard, P.T., Catzeddu, P., Kuipers, O.P., and De Vos, W.M. (2001). Activation of silent *gal* genes in the *lac-gal* regulon of *Streptococcus thermophilus*. *J. Bacteriol.* 183, 1184–1194. doi: 10.1128/JB.183.4.1184-1194.2001.

Vermeulen, N., Ganzle, M.G., and Vogel, R.F. (2007). Glutamine deamidation by cereal-associated lactic acid bacteria. *J. Appl. Microbiol.* 103**,** 1197-1205. doi: 10.1111/j.1365-2672.2007.03333.x.

Vescovo, M., Scolari, G.L., and Bottazzi, V. (1989). Plasmid-encoded ropiness production in *Lactobacillus casei* ssp.*casei*. *Biotechnol. Lett.* 11**,** 709-712. doi: 10.1007/BF01044102.

Vijayakumar, J., Thangavelu, V., and Rajendran, A. (2008). Recent trends in the production, purification and application of lactic acid. *Chem. Biochem. Eng. Q.* 22, 245-264.

Villegas, J.M., Brown, L., Savoy De Giori, G., and Hebert, E.M. (2015). Characterization of the mature cell surface proteinase of *Lactobacillus delbrueckii* subsp. *lactis* CRL 581. *Appl. Microbiol. Biotechnol.* 99**,** 4277-4286. doi: 10.1007/s00253-014-6258-6.

Voelkerding, K.V., Dames, S.A., and Durtschi, J.D. (2009). Next-generation sequencing: From basic research to diagnostics. *Clin. Chem.* 55**,** 641-658. doi: 10.1373/clinchem.2008.112789.

Wang, M., Gao, Z., Zhang, Y., and Pan, L. (2016). Lactic acid bacteria as mucosal delivery vehicles: A realistic therapeutic option. *Appl. Microbiol. Biotechnol.* 100**,** 5691-5701. doi: 10.1007/s00253-016-7557-x.

Warinner, C., Herbig, A., Mann, A., Fellows Yates, J.A., Weiß, C.L., Burbano, H.A., et al. (2017). A robust framework for microbial archaeology. *Annu. Rev. Genomics Hum. Genet.* 18**,** 321-356. doi: 10.1146/annurev-genom-091416-035526.

Weiss, N., Schillinger, U., and Kandler, O. (1983). *Lactobacillus lactis*, *Lactobacillus leichmannii* and *Lactobacillus bulgaricus*, subjective synonyms of *Lactobacillus delbrueckii*, and description of

Lactobacillus delbrueckii subsp. *lactis* comb. nov. and *Lactobacillus delbrueckii* subsp. *bulgaricus* comb. nov. *Syst. Appl. Microbiol.* 4**,** 552-557. doi: 10.1016/s0723-2020(83)80012-5.

Wells, J.M., and Mercenier, A. (2008). Mucosal delivery of therapeutic and prophylactic molecules using lactic acid bacteria. *Nat. Rev. Microbiol.* 6**,** 349-362. doi: 10.1038/nrmicro1840.

Welman, A., and Maddox, I. (2003). Exopolysaccharides from lactic acid bacteria: Perspectives and challenges. *Trends biotechnol.* 21**,** 269-274. doi: 10.1016/S0167-7799(03)00107-0.

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 155-1162. doi: 10.1038/s41587-019-0217-9.

Westra, E.R., Buckling, A., and Fineran, P.C. (2014). CRISPR-Cas systems: Beyond adaptive immunity. *Nat. Rev. Microbiol.* 12**,** 317-326. doi: 10.1038/nrmicro3241.

Whiley, R.A., and Hardie, J.M. (1988). *Streptococcus vestibularis* sp. nov. from the human oral cavity. *Int. J. Syst. Evol. Microbiol.* 38**,** 335-339. doi: 10.1099/00207713-38-4-335.

Wick, R.R., and Holt, K.E. (2019). Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research* 8**,** 2138-2138. doi: 10.12688/f1000research.21782.2.

Williams, A.G., Noble, J., and Banks, J.M. (2001). Catabolism of amino acids by lactic acid bacteria isolated from Cheddar cheese. *Int. Dairy J.* 11**,** 203-215. doi: 10.1016/S0958-6946(01)00050-4.

Williams, A.G., Withers, S.E., and Banks, J.M. (2000). Energy sources of non-starter lactic acid bacteria isolated from Cheddar cheese. *Int. Dairy J.* 10**,** 17-23. doi: 10.1016/S0958-6946(00)00027-3.

Wishart, D.S. (2020). "Genome annotation," in *Bioinformatics,* eds A.D. Baxevanis, G.D. Bader, & D.S. Wishart (Hoboken, NJ: Jonh Wiley & Sons, USA), 117-154.

Wu, Q., and Shah, N.P. (2017). High γ-aminobutyric acid production from lactic acid bacteria: Emphasis on *Lactobacillus brevis* as a functional dairy starter. *Crit. Rev. Food. Sci. Nutr.* 57**,** 3661-3672. doi: 10.1080/10408398.2016.1147418.

Wu, Q., Tun, H. M., Leung, F. C., and Shah, N. P. (2014). Genomic insights into high exopolysaccharide-producing dairy starter bacterium *Streptococcus thermophilus* ASCC 1275. *Sci. Rep.* 4, 4974. doi: 10.1038/srep04974.

Xiong, Z.Q., Kong, L.H., Lai, P.F., Xia, Y.J., Liu, J.C., Li, Q.Y., et al. (2019). Genomic and phenotypic analyses of exopolysaccharide biosynthesis in *Streptococcus thermophilus* S-3. *J. Dairy Sci.* 102**,** 4925-4934. doi: 10.3168/jds.2018-15572.

Xu, Z. (2008) Effect of *Lactobacillus delbrueckii* subsp. *delbrueckii* on the micro flora in broiler gut. *J. Anhui Agric. Univ.* 2008-08.

Xu, Z., Li, S., Gong, G., Liu, Z., Wu, Z., and Ma, C. (2015). Influence of different acidifying strains of *Lactobacillus delbrueckii* subsp. *bulgaricus* on the quality of yoghurt. *Food Sci. Technol. Res.* 21**,** 263-269. doi: 10.3136/fstr.21.263.

Yadav, H., Jain, S., and Sinha, P.R. (2007). Antidiabetic effect of probiotic dahi containing *Lactobacillus acidophilus* and *Lactobacillus casei* in high fructose fed rats. *Nutrition* 23**,** 62-68. doi: 10.1016/j.nut.2006.09.002.

Yamauchi, R., Maguin, E., Horiuchi, H., Hosokawa, M., and Sasaki, Y. (2019). The critical role of urease in yogurt fermentation with various combinations of *Streptococcus thermophilus* and *Lactobacillus delbrueckii* ssp. *bulgaricus*. *J. Dairy Sci.* 102, 1033–1043. doi: 10.3168/jds.2018-15192.

Yang, E., Fan, L., Jiang, Y., Doucette, C., and Fillmore, S. (2012). Antimicrobial activity of bacteriocin-producing lactic acid bacteria isolated from cheeses and yogurts. *AMB Express* 2**,** 48. doi: 10.1186/2191-0855-2-48.

Zacharof, M.P., and Lovitt, R.W. (2012). Bacteriocins produced by lactic acid bacteria a review article. *APCBEE Procedia* 2**,** 50-56. doi: 10.1016/j.apcbee.2012.06.010.

Zannini, E., Waters, D.M., Coffey, A., and Arendt, E.K. (2016). Production, properties, and industrial food application of lactic acid bacteria-derived exopolysaccharides. *Appl. Microbiol. Biotechnol.* 100**,** 1121-1135. doi: 10.1007/s00253-015-7172-2.

Zerbino, D.R. (2010). Using the Velvet *de novo* assembler for short-read sequencing technologies. *Curr. Protoc. Bioinformatics* 31**,** 11.15.11-11.15.12. doi: 10.1002/0471250953.bi1105s31.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* 18**,** 821-829. doi: 10.1101/gr.074492.107.

Zerbino, D.R., Mcewen, G.K., Margulies, E.H., and Birney, E. (2009). Pebble and rock band: Heuristic resolution of repeats and scaffolding in the velvet short-read *de novo* assembler. *PloS One* 4**,** e8407-e8407. doi: 10.1371/journal.pone.0008407.

Zhang, H., Chan, Y., Fan, K., Schmidt, B., and Liu, W. (2018). Fast and efficient short read mapping based on a succinct hash index. *BMC Bioinform.* 19**,** 92-92. doi: 10.1186/s12859-018-2094-5.

Zheng, H.J., Wang, B.F., Zhang, X.L., Han, H., Lu, G., Jin, L., et al. (2008). The complete genome sequence of *Lactobacillus delbrueckii* subsp. *bulgaricus* 2038. *Trends Cell Mol. Biol.* 3**,** 15-30.

Zheng, J., Wittouck, S., Salvetti, E., Franz, C.M.a.P., Harris, H.M.B., Mattarelli, P., et al. (2020). A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int. J. Syst. Evol. Microbiol.* 70**,** 2782-2858. doi: 10.1099/ijsem.0.004107.

Zivkovic, M., Miljkovic, M., Ruas-Madiedo, P., Strahinic, I., Tolinacki, M., Golic, N., et al. (2015). Exopolysaccharide production and ropy phenotype are determined by two gene clusters in putative probiotic strain *Lactobacillus paraplantarum* BGCG11. *Appl. Environ. Microbiol.* 81**,** 1387-1396. doi: 10.1128/aem.03028-14.

Zoumpopoulou, G., Kazou, M., Alexandraki, V., Angelopoulou, A., Papadimitriou, K., Pot, B., et al. (2018a). "Probiotics and prebiotics: An overview on recent trends," in *Probiotics and Prebiotics in Animal Health and Food Safety*, eds D. Di Gioia & B. Biavati B. (Cham: Springer, Switzerland), 1-34.

Zoumpopoulou, G., Tzouvanou, A., Mavrogonatou, E., Alexandraki, V., Georgalaki, M., Anastasiou, R., et al. (2018b). Probiotic features of lactic acid bacteria isolated from a diverse pool of traditional Greek dairy products regarding specific strain-host interactions. *Probiotics Antimicrob. Proteins* 10**,** 313-322. doi: 10.1007/s12602-017-9311-9.

Zuljan, F.A., Mortera, P., Alarcón, S.H., Blancato, V.S., Espariz, M., and Magni, C. (2016). Lactic acid bacteria decarboxylation reactions in cheese. *Int. Dairy J.* 62**,** 53-62. doi: 10.1016/j.idairyj.2016.07.007.

2

Materials and methods

2.1 Bacterial strains and growth conditions

The bacterial strains used in this thesis were isolated from traditionally fermented Greek dairy products. More specifically, *S. thermophilus* ACA-DC 2 and *Lb. delbrueckii* subsp. *bulgaricus* ACA-DC 87 were isolated from traditional Greek yogurt produced from ewe's milk and they were routinely grown in M17 broth (Biokar Diagnostics, Beauvais, France) at 42 ^{o}C for 24 h and MRS broth (Merck, Darmstadt, Germany) at 37 ^{o}C for 24 h, respectively. *Lb. delbrueckii* subsp. *lactis* ACA-DC 178 was isolated from traditional Greek Kasseri cheese made from ewe's milk and was grown in MRS broth at 30 ^{o}C for 24 h. All strains were cultured under microaerophilic conditions.

2.2 Extraction of genomic DNA

For the extraction of genomic DNA of *S. thermophilus* ACA-DC 2, *Lb. bulgaricus* ACA-DC 87 and *Lb. lactis* ACA-DC 178, the protocol developed by Pitcher et al. (1989), was employed. DNA concentration and purity were evaluated with a UV-Vis spectrophotometer (Q5000, Quawell, San Jose, USA). Furthermore, the integrity of the extracted DNA was confirmed electrophoretically in a 0.8% agarose gel.

2.3 PCR amplification of 16S rRNA gene

For the identification of the respective strains at the species level, the 16S rRNA gene was amplified using the two universal oligonucleotide primers A_{For}, 5' -GGA GAG TTA GAT CTT GGC TCA G-3' and C_{Rev}, 5' -AGA AAG GAG GTG ATC CAG CC-3' (Ntougias et al., 2006). The reaction mixture (50 ul) contained 1 µl (50 ng/ul) genomic DNA, 25 ul 2xReadyMix (New England, Biolabs, USA), 1 µl of each primer (10 pmol/ul) and 22 ul ddH_2O. The amplification of genomic DNA was performed using a Mastercycler Gradient thermal cycler (Eppendorf, Germany) with a denaturation step of 2 min at 94 ^{o}C, followed by 30 cycles of 30 s denaturation at 94 ^{o}C, 30 s primer annealing at 56 ^{o}C, 1 min DNA chain extension at 72 ^{o}C and a final 5 min DNA chain extension at 72 ^{o}C. The purification of the PCR products was performed with NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel, Duren, Germany). The purified PCR products were sequenced in VBC-Biotech Service GmbH (Vienna, Austria).

2.4 Genome sequencing and assembly

Whole-genome sequencing of *S. thermophilus* ACA-DC 2 was performed at BaseClear service laboratory for DNA-research (Leiden, The Netherlands) using the PacBio RSII and Illumina HiSeq2500 platforms. The Illumina Casava pipeline v1.8.3 was used for the production of FASTQ sequence files. Quality assessment of the paired-end reads within these files was performed in two consecutively steps. Initially, the data were evaluated by the internal Illumina chastity filter, while the reads with adapters and/or PhiX control signal were warded off using a filtering protocol developed by BaseClear. Subsequently, the quality of the remaining reads was evaluated by the FASTQC quality control tool v0.10.0 and was further improved by trimming off low-quality bases using the bbduk program of the BBMap suite v34.46. Similarly, the continuous long read (CLR) data acquired from the PacBio RSII system were processed and filtered with the SMRT Analysis software suite, based on minimum read length (>50), minimum subread length (>50) and specified read quality (>0.75). The Illumina quality-filtered sequence reads were assembled into contig sequences, employing ABySS v1.5.1, and, subsequently, the PacBio CLR reads were aligned against the pre-assembeld contigs with BLASR method (Chaisson and Tesler, 2012). The information produced by this alignment was further used by the SSPACE-LongRead scaffolder v1.0 to define the orientation, order and distance between the contigs and place them into super-scaffolds (Boetzer and Pirovano, 2014). Finally, GapFiller v1.10 was implemented for closing the gapped regions within the produced super-scaffolds (Boetzer and Pirovano, 2012).

The genomes of *Lb. bulgaricus* ACA-DC 87 and *Lb. lactis* ACA-DC 178 were sequenced with the Illumina HiSeq 2000 platform using three paired-end libraries (500-bp, 2,000-bp, and 6,000-bp) at the Beijing Genomics Institute (BGI Co., Ltd., Hong Kong). After filtering, reads were assembled with SOAPdenovo v1.05 (*Lb. bulgaricus* ACA-DC 87) and v2.04 (*Lb. lactis* ACA-DC 178) (Li et al., 2008; Li et al., 2009). For *Lb. bulgaricus* ACA-DC 87 assembly accuracy was evaluated through whole-genome alignment of the ACA-DC 87 genome against a reference genome, namely that of *Lb. bulgaricus* ATCC 11842 (van de Guchte et al., 2006), using ProgressiveMauve (Darling et al., 2010). The accuracy of the hybrid assembly for *Lb. lactis* ACA-DC 178 was evaluated through the chromosomal optical map of the strain, which was generated at Microbion SRL

(Verona, Italy; Latreille et al., 2007). The alignment of the assembly against the optical map was created with MapSolver (OpGen Technologies, Inc., Madison, WI).

2.5 Structural annotation

Structural annotation of the three LAB genomes was performed by employing three independent *ab initio* online resources, namely FGENESB (Solovyev and Salamov, 2011) MetaGeneAnnotator (Noguchi et al., 2008) and Prodigal (Hyatt et al., 2010). This multi-tool strategy was followed aiming at comparing and verifying the results, i.e. the coordinates of putative genes produced by the different methods. Additionally, genome annotation, including prediction of rRNA and tRNA genes, was executed with RAST (Aziz et al., 2008), while the GenePRIMP evidence-based method was used for the identification of putative pseudogenes (Pati et al., 2010). The quality of the annotation was further assessed and improved by manual curation of the generated data using BLAST suite for sequence similarity searches (Altschul et al., 1990) and Artemis genome viewer/browser tool (Carver et al., 2012).

2.6 Functional annotation

Visualization of the genomic features identified for each of the three LAB strains, including GC content and CG skew, was achieved through the generation of circular genome maps. More specifically, for *S. thermophilus* ACA-DC 2 genome the CGView comparison tool was used (Grant and Stothard, 2008), while for *Lb. bulgaricus* ACA-DC 87 and *Lb. lactis* ACA-DC 178 genomes the DNAPlotter software was employed (Carver et al., 2008). Further bioinformatics analysis included COG functional annotation of the predicted proteins with WebMGA server (strains ACA-DC 2 and ACA-DC 87; Wu et al., 2011) and the EggNog v4.5 (strains ACA-DC 2 and ACA-DC 178; Huerta-Cepas et al., 2016; Huerta-Cepas et al., 2017), identification of genes with Pfam domains according to the Pfam protein families database (strains ACA-DC 2 and ACA-DC 87; Finn et al., 2016), prediction of putative CRISPR arrays employing the CRISPRFinder web tool (Grissa et al., 2007), identification of GIs through the IslandViewer 3 (strain ACA-DC 2; Dhillon et al., 2015) and IslandViewer 4 (strains ACA-DC 87 and ACA-DC 178; Bertelli et al., 2017), determination of R-M systems using the REBASE database (Roberts et al., 2010; 2015), and finally mining of putative

antimicrobial peptides with BAGEL3 (strain ACA-DC 2; van Heel et al., 2013) and BAGEL4 (strains ACA-DC 87 and ACA-DC 178; van Heel et al, 2018). Furthermore, for *S. thermophilus* ACA-DC 2 the Phobius web server was implemented for the identification of genes with transmembrane helices and genes with signal peptides (Kall et al., 2007), while potential pathogenic features were detected using the MP3 tool (Gupta et al., 2014).

2.7 Comparative and evolutionary genomics

S. thermophilus comparative genomics analysis was performed employing two different sets of strains, including 23 (see Chapter 3) and 15 (see Chapter 4) strains, respectively. For the set of 15 *S. thermophilus* strains, phylogenetic analysis based on 16S rRNA gene was performed. 16S rRNA gene sequences of strains belonging to diverse *Streptococcus* species were also included in the analysis. The sequences were aligned using MUSCLE (Edgar, 2004) and the phylogenetic tree was constructed by the maximum likelihood method using the MEGA7 software (Kumar et al., 2016) using the Tamura-Nei substitution model (Tamura and Nei, 1993). *Lactococcus lactis* subsp. *lactis* NCDO 604^T served as the outgroup. Bootstrap values derived after 1,000 replicates. Additionally, whole genome phylogenetic analysis was executed with the EDGAR server (Blom et al., 2016). The resulting phylogenetic tree was presented as a cladogram ignoring branch length. *S. salivarius* NCTC 8616 and *L. lactis subsp. cremoris* MG1363 were used as outgroups. Further comparative analysis among KLDS 3.1003, JIM 8232 and ACA-DC 2 strains included Venn diagram analysis, as implemented in EDGAR (Blom et al., 2016) and whole genome alignment with Circoletto (Darzentas et al., 2010). Finally, the synteny between the CRISPR loci of ACA-DC 2 and LMD-9 strains was visualized using the KODON software (Applied Maths NV, Sint-Martens-Latem, Belgium).

The analysis of the 23 *S. thermophilus* strains was based on numerous bioinformatics tools. Whole genome alignment was produced by ProgressiveMauve (Darling et al., 2010), while the identified chromosomal inversions in strains EPS, MN-BM-A01, and MN-ZLW-002 were assessed with GenSkew online tool (http://genskew.csb.univie.ac.at/). The bacterial pan genome analysis (BPGA) pipeline v.1.3 (Chaudhari et al., 2016) was employed for estimating the pan/core genome of the species, using USEARCH v.9.2.64 for gene families clustering (Edgar, 2010). Certain

parameters were optimally adjusted to facilitate the analysis, including sequence identity cut-off (60%) and random permutations of genomes (20). The latter was implemented for avoiding any bias within the successive addition of new genomes. This pipeline was also used for COG functional classification of the proteins assigned in the core, accessory and unique gene families (Chaudhari et al., 2016). The respective proteins were also analyzed for COG categories with the eggNOG v4.5 (Huerta-Cepas et al., 2016, Huerta-Cepas et al., 2017). The EDGAR software was employed for analysis of orthologs and core genome phylogenetic analysis among *S. thermophilus* strains (Blom et al., 2016). More specifically, the alignments of the core gene sets were executed with MUSCLE and subsequently concatenated to one complete core alignment, which was used to generate the phylogenetic tree by the neighbor-joining method as implemented in the Phylogeny Inference Package (PHYLIP) package. The topology of the consensus tree was confirmed by 100 bootstrap iterations. The relatedness among *S. thermophilus* strains was investigated also by the EDGAR software, through the construction of an average nucleotide identity (ANI) heat map. The computed phylogenetic distance values (Goris et al., 2007; Richter and Rosselló-Móra, 2009) were arranged in an ANI matrix, clustered in consistence with their distance patterns and visualized as a color-coded heatmap (Blom et al., 2016). For assessing the statistical differences in genome size, the Mann-Whitney U Test for $p < 0.05$ was employed, while the respective differentiation was visualized with BoxPlot Generator (https://plot.ly). The microbial genomes atlas (MiGA) webserver was used for evaluating the quality of genome assemblies (Rodriguez-R et al., 2018). Heatmaps for COG frequency and presence/absence of accessory genes were generated with the RStudio by means of the heatmap.2 function included in the Gplots package (http://www.rstudio.org). Assignment of K number to *S. thermophilus* protein-coding sequences was performed with KEGG orthology and links annotation (KOALA) (Kanehisa et al., 2016b), and the resulting KEGG orthology (KO) annotations were also processed with the KEGG Mapper tools (Kanehisa et al., 2016a). The architecture and the gene content of the EPS loci were compared using the Easyfig comparison visualizer (Sullivan et al., 2011). The peptide and amino acid transporters were identified using the TransportDB database (Elbourne et al., 2017). Putative prophage sequences were determined with the PHAge Search Tool-Enhanced Release (PHASTER) web server (Arndt et al.,

2016). CRISPRFinder (Grissa et al., 2007) and CD-HIT Suite (Huang et al., 2010) were employed for the identification of CRISPR arrays and the comparison of the predicted spacers, respectively. Finally, the REBASE database was used for identification of R-M system components (Roberts et al., 2010; 2015), while IslandViewer 4 was utilized in the detection of GIs (Bertelli et al., 2017).

2.8 References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215**,** 403-410. doi: 10.1016/S0022-2836(05)80360-2.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44**,** W16-21. doi: 10.1093/nar/gkw387.

Aziz, R.K., Bartels, D., Best, A.A., Dejongh, M., Disz, T., Edwards, R.A., et al. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genom.* 9**,** 75-75. doi: 10.1186/1471-2164-9-75.

Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing Group, Lau, B. Y., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343.

Blom, J., Kreis, J., Spanig, S., Juhre, T., Bertelli, C., Ernst, C., et al. (2016). EDGAR 2.0: An enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* 44**,** W22-28. doi: 10.1093/nar/gkw255.

Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biol.* 13**,** R56-R56. doi: 10.1186/gb-2012-13-6-r56.

Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinform.* 15**,** 211. doi: 10.1186/1471-2105-15-211.

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., and Mcquillan, J.A. (2012). Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28**,** 464-469. doi: 10.1093/bioinformatics/btr703.

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2008). DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics* 25**,** 119-120. doi: 10.1093/bioinformatics/btn578.

Chaisson, M.J., and Tesler, G. (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinform.* 13**,** 238. doi: 10.1186/1471-2105-13-238.

Chaudhari, N.M., Gupta, V.K., and Dutta, C. (2016). BPGA-an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* 6**,** 24373. doi: 10.1038/srep24373.

Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5**,** e11147. doi: 10.1371/journal.pone.0011147.

Darzentas, N. (2010). Circoletto: Visualizing sequence similarity with Circos. *Bioinformatics* 26**,** 2620-2621. doi: 10.1093/bioinformatics/btq484.

Dhillon, B.K., Laird, M.R., Shay, J.A., Winsor, G.L., Lo, R., Nizam, F., et al. (2015). IslandViewer 3: More flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res.* 43**,** W104-108. doi: 10.1093/nar/gkv401.

Edgar, R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792-7. doi: 10.1093/nar/gkh340.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26**,** 2460-2461. doi: 10.1093/bioinformatics/btq461.

Elbourne, L.D.H., Tetu, S.G., Hassan, K.A., and Paulsen, I.T. (2017). TransportDB 2.0: A database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.* 45**,** D320-D324. doi: 10.1093/nar/gkw1068.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., et al. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* 44**,** D279-285. doi: 10.1093/nar/gkv1344.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57**,** 81-91. doi: 10.1099/ijs.0.64483-0.

Grant, J.R., and Stothard, P. (2008). The CGView Server: A comparative genomics tool for circular genomes. *Nucleic Acids Res.* 36**,** W181-184. doi: 10.1093/nar/gkn179.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35**,** W52-57. doi: 10.1093/nar/gkm360.

Gupta, A., Kapil, R., Dhakan, D.B., and Sharma, V.K. (2014). MP3: A software tool for the prediction of pathogenic proteins in genomic and metagenomic data. *PLoS One* 9**,** e93907. doi: 10.1371/journal.pone.0093907.

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT suite: A web server for clustering and comparing biological sequences. *Bioinformatics* 26**,** 680-682. doi: 10.1093/bioinformatics/btq003.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., Von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34**,** 2115-2122. doi: 10.1093/molbev/msx148.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., et al. (2016). eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44**,** D286-293. doi: 10.1093/nar/gkv1248.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11**,** 119. doi: 10.1186/1471-2105-11-119.

Kall, L., Krogh, A., and Sonnhammer, E.L. (2007). Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. *Nucleic Acids Res.* 35**,** W429-432. doi: 10.1093/nar/gkm256.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016a). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44**,** D457-D462. doi: 10.1093/nar/gkv1070.

Kanehisa, M., Sato, Y., and Morishima, K. (2016b). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428**,** 726-731. doi: 10.1016/j.jmb.2015.11.006.

Kumar, S., Stecher, G., Tamura, K. (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870-4. doi: 10.1093/molbev/msw054.

Latreille, P., Norton, S., Goldman, B.S., Henkhaus, J., Miller, N., Barbazuk, B., et al. (2007). Optical mapping as a routine tool for bacterial genome sequence finishing. *BMC Genom.* 8**,** 321-321. doi: 10.1186/1471-2164-8-321.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008). SOAP: Short oligonucleotide alignment program. *Bioinformatics* 24**,** 713-714. doi: 10.1093/bioinformatics/btn025.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., et al. (2009). SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* 25**,** 1966-1967. doi: 10.1093/bioinformatics/btp336.

Noguchi, H., Taniguchi, T., and Itoh, T. (2008). MetaGeneAnnotator: Detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* 15**,** 387-396. doi: 10.1093/dnares/dsn027.

Ntougias, S., Zervakis, G.I., Ehaliotis, C., Kavroulakis, N., Papadopoulou, K.K. (2006). Ecophysiology and molecular phylogeny of bacteria isolated from alkaline two-phase olive mill wastes. *Res. Microbiol*. 157, 376-385. doi: 10.1016/j.resmic.2005.09.010.

Pati, A., Ivanova, N.N., Mikhailova, N., Ovchinnikova, G., Hooper, S.D., Lykidis, A., et al. (2010). GenePRIMP: A gene prediction improvement pipeline for prokaryotic genomes. *Nat. Methods* 7**,** 455-457. doi: 10.1038/nmeth.1457.

Pitcher, D.G., Saunders, N.A., and Owen, R.J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Lett. Appl. Microbiol.* 8**,** 151-156. doi: 10.1111/j.1472-765X.1989.tb00262.x.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106**,** 19126-19131. doi: 10.1073/pnas.0906412106.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2010). REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.* 38**,** D234-D236. doi: 10.1093/nar/gkp874.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.* 43**,** D298-D299. doi: 10.1093/nar/gku1046.

Rodriguez-R, L.M., Gunturu, S., Harvey, W.T., Rosselló-Móra, R., Tiedje, J.M., Cole, J.R., et al. (2018). The Microbial Genomes Atlas (MiGA) webserver: Taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.* 46**,** W282-W288. doi: 10.1093/nar/gky467.

Solovyev, V., and Salamov, A. (2011). "Automatic annotation of microbial genomes and metagenomic sequences," in *Metagenomics and its applications in agriculture, biomedicine and environmental studies,* ed. R.W. Li (New York, NY: Nova Science Publishers), 67-78.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011). Easyfig: A genome comparison visualizer. *Bioinformatics* 27**,** 1009-1010. doi: 10.1093/bioinformatics/btr039.

Tamura, K., and Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512-26. doi: 10.1093/oxfordjournals.molbev.a040023.

van De Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., et al. (2006). The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc. Natl. Acad. Sci. U.S.A.* 103**,** 9274-9279. doi: 10.1073/pnas.0603024103.

van Heel, A.J., De Jong, A., Montalban-Lopez, M., Kok, J., and Kuipers, O.P. (2013). BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res.* 41**,** W448-453. doi: 10.1093/nar/gkt391.

van heel, A.J., De jong, A., Song, C., Viel, J.H., Kok, J., and Kuipers, O.P. (2018). BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* 46**,** W278-W281. doi: 10.1093/nar/gky383.

Wu, S., Zhu, Z., Fu, L., Niu, B., and Li, W. (2011). WebMGA: A customizable web server for fast metagenomic sequence analysis. *BMC Genom.* 12**,** 444. doi: 10.1186/1471-2164-12-444.

3

Comparative genomics of *Streptococcus thermophilus* support important traits concerning the evolution, biology and technological properties of the species

frontiers in Microbiology

ORIGINAL RESEARCH
published: 20 December 2019
doi: 10.3389/fmicb.2019.02916

Comparative Genomics of *Streptococcus thermophilus* Support Important Traits Concerning the Evolution, Biology and Technological Properties of the Species

Voula Alexandraki[1], Maria Kazou[1], Jochen Blom[2], Bruno Pot[3], Konstantinos Papadimitriou[1†] and Effie Tsakalidou[1]*

[1] Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Athens, Greece, [2] Bioinformatics and Systems Biology, Justus Liebig University Giessen, Giessen, Germany, [3] Research Group of Industrial Microbiology and Food Biotechnology (IMDO), Department of Bioengineering Sciences (DBIT), Vrije Universiteit Brussel, Brussels, Belgium

Streptococcus thermophilus is a major starter for the dairy industry with great economic importance. In this study we analyzed 23 fully sequenced genomes of *S. thermophilus* to highlight novel aspects of the evolution, biology and technological properties of this species. Pan/core genome analysis revealed that the species has an important number of conserved genes and that the pan genome is probably going to be closed soon. According to whole genome phylogeny and average nucleotide identity (ANI) analysis, most *S. thermophilus* strains were grouped in two major clusters (i.e., clusters A and B). More specifically, cluster A includes strains with chromosomes above 1.83 Mbp, while cluster B includes chromosomes below this threshold. This observation suggests that strains belonging to the two clusters may be differentiated by gene gain or gene loss events. Furthermore, certain strains of cluster A could be further subdivided in subgroups, i.e., subgroup I (ASCC 1275, DGCC 7710, KLDS SM, MN-BM-A02, and ND07), II (MN-BM-A01 and MN-ZLW-002), III (LMD-9 and SMQ-301), and IV (APC151 and ND03). In cluster B certain strains formed one distinct subgroup, i.e., subgroup I (CNRZ1066, CS8, EPS, and S9). Clusters and subgroups observed for *S. thermophilus* indicate the existence of lineages within the species, an observation which was further supported to a variable degree by the distribution and/or the architecture of several genomic traits. These would include exopolysaccharide (EPS) gene clusters, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)-CRISPR associated (Cas) systems, as well as restriction-modification (R-M) systems and genomic islands (GIs). Of note, the histidine biosynthetic cluster was found present in all cluster A strains (plus strain NCTC12958^T) but was absent from all strains in cluster B. Other loci related to lactose/galactose catabolism and urea metabolism, aminopeptidases, the majority of amino acid and peptide transporters, as well as amino acid biosynthetic pathways

OPEN ACCESS

Edited by:
Nikos Kyrpides, Lawrence Berkeley National Laboratory, United States

Reviewed by:
Stefano Campanaro, University of Padua, Italy
Anastasia Chasapi, Centre for Research & Technology Hellas, Greece

****Correspondence:***
Konstantinos Papadimitriou
kpapadimitriou@aua.gr

†Present address:
Konstantinos Papadimitriou, Department of Food Science and Technology, Faculty of Agriculture and Foods, University of Peloponnese, Kalamata, Greece

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *09 August 2019*
Accepted: *03 December 2019*
Published: *20 December 2019*

Citation:
Alexandraki V, Kazou M, Blom J, Pot B, Papadimitriou K and Tsakalidou E (2019) Comparative Genomics of Streptococcus thermophilus Support Important Traits Concerning the Evolution, Biology and Technological Properties of the Species. Front. Microbiol. 10:2916. doi: 10.3389/fmicb.2019.02916

were found to be conserved in all strains suggesting their central role for the species. Our study highlights the necessity of sequencing and analyzing more *S. thermophilus* complete genomes to further elucidate important aspects of strain diversity within this starter culture that may be related to its application in the dairy industry.

Keywords: lineage, horizontal gene transfer, genomic islands, milk, yogurt, cheese, pan genome, CRISPR

INTRODUCTION

Lactic acid bacteria (LAB) include several species, which are extensively used as starters in dairy fermentations (Kongo, 2013). Among them, *Streptococcus thermophilus* constitutes a major starter for the dairy industry. It is primarily used in the production of yogurt, alongside with *Lactobacillus delbrueckii* subsp. *bulgaricus*, but also in the production of several cheese varieties, such as Feta and Mozzarella (Purwandari et al., 2007; Rantsiou et al., 2008; Anbukkarasi et al., 2013). *S. thermophilus* is the only species which was granted the generally recognized as safe (GRAS) status according to the Food and Drug Administration [FDA], 2007 and the qualified presumption of safety (QPS) status according to the European Food Safety Authority [EFSA], 2007 within the *Streptococcus* genus, which consists mainly of commensals and pathogenic species. As it is attested by the large number of pseudogenes identified in the genomes of the *S. thermophilus* strains sequenced so far, the species has undergone significant genome decay probably due to its adaptation to the dairy environment, which is particularly rich in nutrients (Bolotin et al., 2004; Hols et al., 2005; Goh et al., 2011). The regressive evolution of the species has led to genome reduction and simplification of its metabolism (Mayo et al., 2008). The latter is reflected in the deterioration of genes involved, among others, in sugar utilization. *S. thermophilus* has also lost typical streptococcal pathogenic features presumably through strain selection during domestication toward a starter culture (Bolotin et al., 2004; Hols et al., 2005; Goh et al., 2011; Papadimitriou et al., 2015b). Furthermore, the protocooperation with *L. bulgaricus* during the production of yogurt has further shaped the metabolic properties of *S. thermophilus* toward this symbiotic relationship (Mayo et al., 2008).

Typical technological features of *S. thermophilus*, such as milk acidification, lactose and galactose utilization, proteolytic activity and exopolysaccharide (EPS) production, contribute in shaping the organoleptic characteristics of the final products (Cui et al., 2016). In addition, the stress responses of the species define its performance under the unfavorable conditions prevailing during food production (Zotta et al., 2008; Cui et al., 2016). *S. thermophilus* also carries Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)-CRISPR associated (Cas) (CRISPR-Cas) and restriction-modification (R-M) systems, which may contribute to competitiveness in microbial food ecosystems and resistance against bacteriophages and other parasitic DNA (Horvath and Barrangou, 2010; Dupuis et al., 2013). Moreover, genes in genomic islands (GIs), which have been acquired most probably through horizontal gene transfer (HGT) events, may ascribe a number of adaptive traits to *S. thermophilus* and could be related to technological characteristics, such as EPS production, bacteriocin biosynthesis, and protocooperation (Liu et al., 2009; Eng et al., 2011).

Another topic that has attracted some attention concerning *S. thermophilus* was the biodiversity of strains within the species. Original studies used typing techniques like random amplification of polymorphic DNA-PCR (RAPD-PCR) or pulsed-field gel electrophoresis (PFGE), while more recent ones used multilocus sequence typing (MLST) (Moschetti et al., 1998; Giraffa et al., 2001; Mora et al., 2002; Ercolini et al., 2005; Delorme et al., 2010, 2017; Yu et al., 2015). Application of MLST in *S. thermophilus* had to be optimized to increase discriminating power, given the fact that the species may exhibit limited genetic variability (Delorme et al., 2017). In this study the authors reported 116 sequence types and the existence of groups of strains based on phylogenetic analysis of concatenated sequences of housekeeping genes. Additional analysis revealed clustering of strains based on core genome and CRISPR spacer analysis of 25 sequenced strains (both complete and partial). The authors reported that the clustering based on MLST and whole genome analysis was in agreement but differed from that of CRISPR analysis. With the MLST scheme developed and the wide sample of *S. thermophilus* strains (n = 178), it was feasible to detect relationship between strains and geographic location.

Furthermore, due to the economic importance of *S. thermophilus* as a starter, a number of groundbreaking studies have been conducted in an attempt to elucidate the genetic basis behind the physiological and the metabolic properties of the species, which define its technological and probiotic potential. Comparative genomics of *S. thermophilus* was carried out early on and provided significant information about its adaptation to the milk environment and technological traits (Bolotin et al., 2004; Hols et al., 2005; Goh et al., 2011). However, these studies relied only on a limited number of genome sequences. The current accumulation of completely sequenced *S. thermophilus* genomes can increase the predictive

Abbreviations: ABC, ATP-binding cassette; ANI, average nucleotide identity; APC, amino acid-polyamine-organocation; blp, bacteriocin-like peptide; BPGA, bacterial pan genome analysis; COG, clusters of orthologous groups; CRISPR-Cas, clustered regularly interspaced short palindromic repeats-CRISPR associated; DRs, direct repeats; EFSA, European Food Safety Authority; EPS, exopolysaccharide; FDA, Food and Drug Administration; GABA, gamma-aminobutyric acid; GIs, genomic islands; GIT, gastrointestinal tract; GRAS, generally recognized as safe; GSH, glutathione; HGT, horizontal gene transfer; ISs, insertion sequences; KEGG, Kyoto Encyclopedia of Genes and Genomes; KOALA, KEGG orthology and links annotation; LAB, lactic acid bacteria; LCBs, local collinear blocks; MiGA, microbial genomes atlas; MLST, multilocus sequence typing; ORFs, open reading frames; PFGE, pulsed-field gel electrophoresis; PFL, pyruvate formate lyase; PFLA, pyruvate formate-lyase activating; PGAP, prokaryotic genome annotation pipeline; QPS, qualified presumption of safety; RAPD, random amplification of polymorphic DNA; RAST, rapid annotation using subsystem technology; R-M, restriction-modification; ROS, reactive oxygen species; SBSEC, *Streptococcus bovis*/*Streptococcus equinus* complex.

power of comparative analysis and enhance the interpretation of the acquired data about the genome architecture, functionality and evolution. Furthermore, the advancement of bioinformatics tools and the demand of the dairy industry for novel starter strains render an updated analysis of the species essential. In the present study, the results of an in depth analysis of 23 complete *S. thermophil*us genomes are presented, focusing on main technological features of the species.

MATERIALS AND METHODS

Strains

The 23 *S. thermophilus* genomes designated as "complete" up to RefSeq release 88, were selected for analysis in this study (**Table 1**). The majority of *S. thermophilus* strains have been isolated from yogurt (strains LMG 18311, CNRZ1066, LMD-9, MN-ZLW-002, MN-BM-A01, KLDS SM, KLDS 3.1003, and ACA-DC 2) and milk (strains JIM 8232, SMQ-301, ND03, ND07, B59671, EPS, GABA, and NCTC12958^T). Furthermore, three isolates, namely strains S9, MN-BM-A02, and CS8, derived from traditional Chinese dairy products. More specifically, MN-BM-A02 was isolated from Fan, a traditional Chinese cheese-like product, while CS8 from Rubing, a Chinese fresh goat milk cheese. Finally, strains APC151 and ST3 were isolated from fish intestine and commercial dietary supplements, respectively.

Comparative and Evolutionary Genomics

ProgressiveMauve was used for the whole genome alignment of the 23 *S. thermophilus* strains analyzed in this study (Darling et al., 2010). GenSkew online application was employed in the evaluation of the chromosomal inversions in strains EPS, MN-BM-A01, and MN-ZLW-002[1]. The pan/core genome analysis was performed with the bacterial pan genome analysis (BPGA) pipeline v.1.3 using USEARCH v.9.2.64 for clustering gene families (Edgar, 2010) with a 60% sequence identity cut-off and 20 random permutations of genomes to avoid any bias in the sequential addition of new genomes. The protein coding sequences assigned in the core, accessory and unique gene families were further analyzed for clusters of orthologous groups (COG) categories within the BPGA pipeline (Chaudhari et al., 2016). Alternatively, protein coding sequences of *S. thermophilus* strains were also analyzed for COG categories with the eggnog-mapper based on eggNOG v.4.5 orthology database, as highlighted in the text (Huerta-Cepas et al., 2016, 2017). The EDGAR tool was also employed to assist analysis of orthologs whenever necessary, as well as for core genome phylogenetic analysis among *S. thermophilus* strains (Blom et al., 2016). For the latter, the alignments of the core gene sets were executed with MUSCLE and concatenated to one complete core alignment, which was used to generate the phylogenetic tree by the neighbor-joining method as implemented in the PHYLIP package. The consensus tree topology was verified by 100 bootstrap iterations. The EDGAR software was also exploited for the investigation of the relatedness among *S. thermophilus* strains through the construction of average nucleotide identity (ANI) heat map. The ANI values were computed as described by Goris et al. (2007) and as implemented in the JSpecies package (Richter and Rossello-Mora, 2009). The resulting phylogenetic distance values were arranged in an ANI matrix, clustered according to their distance patterns and visualized as a color-coded heatmap, with dark and light orange for high and low similarity regions, respectively. Box Plot Generator was employed for the visualization of genome size differences between the two clusters of the *S. thermophilus* strains[2]. Statistical differences in genome size were accessed with the Mann–Whitney U Test for $p < 0.05$. The quality of the genome assemblies was evaluated with the microbial genomes atlas (MiGA) webserver (Rodriguez-R et al., 2018). The COG frequency and the accessory genes presence/absence heatmaps were generated with the RStudio using the heatmap.2 function included in the Gplots package[3]. Kyoto encyclopedia of genes and genomes (KEGG) orthology and links Annotation (KOALA) was employed for K number assignment to *S. thermophilus* protein coding sequences (Kanehisa et al., 2016b), while KEGG Mapper tools were exploited for further processing of KO annotations (Kanehisa et al., 2016a). The PHASTER web server was used for the identification of putative prophages (Arndt et al., 2016). The comparison of the EPS gene clusters was performed with the Easyfig tool (Sullivan et al., 2011). The transporters were determined using the TransportDB database (Elbourne et al., 2017). The CRISPRs were identified with CRISPRFinder web tool (Grissa et al., 2007), while comparison of the predicted spacers was performed with CD-HIT Suite (Huang et al., 2010). The REBASE database was used for verifying the R-M systems (Roberts et al., 2015). Finally, the GIs were obtained through the IslandViewer 4 web-based resource (Bertelli et al., 2017). For our analysis, GIs characterized as integrated by the IslandViewer tool were analyzed.

RESULTS AND DISCUSSION

General Genomic Features

The general genome features of the 23 *S. thermophilus* strains used in this study are presented in **Table 2**. The chromosome length of the strains ranges between 1.73 and 2.10 Mbp, with an average of 1.85 Mbp, while the % GC content is around 39.0. The number of genes varied between 1,847 and 2,237 including protein coding sequences that varied between 1,555 and 1,854. The percentage of pseudogenes ranged between 9.64 and 13.97%. These variations in genome size, gene and pseudogene content indicate important differences in both gene gain and gene loss events during the evolution of the different strains. It has been previously reported that *S. thermophilus* owns some of the smallest genomes within streptococci while *Streptococcus salivarius* some of the largest (Delorme et al., 2015). Based on the complete genome sequences within the salivarius group we found that the percentage of pseudogenes of *S. salivarius* (12 complete genomes) may reach up to 4%

[1]http://genskew.csb.univie.ac.at/

[2]https://plot.ly

[3]http://www.rstudio.org

TABLE 1 | *Streptococcus thermophilus* strains with complete genomes analyzed in this study.

Strain	GenBank accession	Isolation source	Sequencing technology	References
LMG 18311	NC_006448	Commercial yogurt	Random shotgun sequencing	Bolotin et al., 2004
CNRZ1066	NC_006449	Commercial Yogurt	Random shotgun sequencing	Bolotin et al., 2004
LMD-9	NC_008532	Yogurt	Whole-genome shotgun sequencing	Makarova et al., 2006
ND03	NC_017563	Naturally fermented yak milk	454; Solexa	Sun et al., 2011
JIM 8232	NC_017581	Raw milk	SOLiD; Sanger	Delorme et al., 2011
MN-ZLW-002	NC_017927	Traditional yogurt block	454; Solexa	Kang et al., 2012
ASCC 1275	NZ_CP006819	–	454	Wu et al., 2014
SMQ-301	NZ_CP011217	Milk	Illumina; PacBio	Labrie et al., 2015
MN-BM-A02	NZ_CP010999	Dairy fan	454 GS FLX	Shi et al., 2015
MN-BM-A01	NZ_CP012588	Traditional yogurt block	PacBio RS	Bai et al., 2016
KLDS 3.1003	NZ_CP016877	Traditional yogurt	Illumina	Evivie et al., 2017
ACA-DC 2	NZ_LT604076	Traditional yogurt	Illumina HiSeq2500; PacBio RSII	Alexandraki et al., 2017
APC151	NZ_CP019935	Fish intestine	PacBio RS	Linares et al., 2016, 2017
B59671	NZ_CP022547	Raw milk	PacBio RS	Renye et al., 2017
KLDS SM	NZ_CP016026	Traditional yogurt	Illumina	Li et al., 2018
DGCC 7710	NZ_CP025216	Dairy culture	Illumina MiSeq; PacBio RS	Hatmaker et al., 2018
S9	NZ_CP013939	Traditional dairy	PacBio	–
CS8	NZ_CP016439	Rubing	PacBio	–
ND07	NZ_CP016394	Naturally fermented yak milk	PacBio RSII	–
EPS	NZ_CP025400	Milk	PacBio RS	–
GABA	NZ_CP025399	Milk	PacBio RS	–
ST3	NZ_CP017064	Commercial dietary supplements	PacBio RS	–
NCTC12958^T	NZ_LS483339	Milk	–	–

while the percentage of pseudogenes of *Streptococcus vestibularis* NCTC12167, the only strain with a complete genome, was around 8%. These findings suggest a variable degree of evolution through genome decay within the group. Beyond the salivarius group, high percentages of pseudogenes have also been reported for *Streptococcus macedonicus* and *Streptococcus infantarius* that are also associated with the dairy environment (Jans et al., 2013a; Papadimitriou et al., 2014). A high number of pseudogenes has also been reported for certain strains of *Streptococcus pneumoniae* (see for example the studies by Junges et al., 2019; Scott et al., 2019). Interestingly, extensive genome decay seems to be compatible with adaptation in milk (Bolotin et al., 2004; Hols et al., 2005; Jans et al., 2013a; Papadimitriou et al., 2014) or a pathogenic lifestyle (Lerat and Ochman, 2005). Obviously more research is needed to appreciate the strains/species within streptococci that have evolved through reductive processes and to test whether this evolution path can be correlated with the niches they occupy.

Fourteen out of 23 strains carry 18 rRNA genes and the rest carry 15. Interestingly, strains with 18 rRNA genes also own a higher number of tRNA genes (ranging from 67 to 69) compared to strains with 15 rRNAs which own fewer tRNA genes (ranging from 55 to 57). A general comment that can be made about this difference is that strains with a higher number of rRNA and tRNA genes could potentially exhibit a higher growth/metabolic rate (Wassenaar and Lukjancenko, 2014).

Comparison of the chromosomal architecture of the 23 *S. thermophilus* strains was performed through full-length sequence alignments (**Supplementary Figure S1**). All strains were synchronized from the *dnaA* so as to simplify the alignment. Analysis revealed a high degree of conservation among different strains. However, strain-specific differences could also be detected. More specifically, low similarity regions, represented as white regions inside the local collinear blocks (LCBs), were found in all strains. Furthermore, many unique regions, represented as blank spaces between the LCBs, were also identified in all strains. In strain EPS a large inversion (1.47 Mbp) was present, while in strains MN-BM-A01 and MN-ZLW-002, a ∼300 kbp inverted region was identified between coordinates 768,310–1,068,868 and 740,416–1,040,999 bp, respectively. These inversions could be either genuine or could be ascribed to assembly artifacts. If the first is true, our observations may correspond to an inversion around the origin of replication for strain EPS, or to an inversion around the terminus of replication for strains MN-BM-A01 and MN-ZLW-002. Such inversions have been described before for bacterial genomes as part of their evolution (Eisen et al., 2000; Darling et al., 2008; Repar and Warnecke, 2017).

Pan/Core Genome Analysis and Phylogenomics

The pan genome of the 23 *S. thermophilus* strains contains a total number of 2,516 genes, including 1,082 and 997 genes in the core and accessory genomes, respectively (**Figure 1A**). The number of genes in the accessory genome of each strain varied between 432 and 568 and a total of 437 unique genes (singletons) were identified in 14 strains (**Supplementary Table S1** and

TABLE 2 | General genome features of *S. thermophilus* strains with complete genomes analyzed in this study.

Strain	Genome size (bp)	GC (%)	Genes	Proteins	rRNA	tRNA	Pseudogenes (% of pseudogenes)	Predicted essential genes[1]	Genome completeness (%)[2]	Corrected genome completeness (%)[3]
NCTC12958^T	2,102,271	39.0	2,237	1,854	15	56	308 (13.77)	106	95.5	100.0
JIM 8232	1,929,905	38.9	2,033	1,748	18	67	196 (9.64)	104	93.7	98.1
KLDS 3.1003	1,899,956	38.9	2,037	1,676	18	68	271 (13.30)	105	94.6	99.1
MN-BM-A01	1,876,516	39.1	2,023	1,661	18	67	273 (13.49)	104	93.7	98.1
ND07	1,869,510	39.0	1,996	1,684	15	57	236 (11.82)	105	94.6	99.1
ST3	1,865,056	39.0	1,982	1,638	18	69	253 (12.76)	106	95.5	100.0
SMQ-301	1,861,792	39.1	1,993	1,684	18	67	220 (11.04)	106	95.5	100.0
GABA	1,857,468	39.1	1,952	1,621	18	68	241 (12.35)	106	95.5	100.0
KLDS SM	1,856,787	39.1	1,984	1,671	18	67	224 (11.29)	106	95.5	100.0
LMD-9	1,856,368	39.1	1,993	1,674	18	67	230 (11.54)	105	94.6	99.1
DGCC 7710	1,851,207	39.0	1,962	1,657	15	56	230 (11.72)	106	95.5	100.0
MN-BM-A02	1,850,434	39.0	1,977	1,677	15	57	224 (11.33)	106	95.5	100.0
MN-ZLW-002	1,848,520	39.1	1,982	1,695	15	57	211 (10.65)	105	94.6	99.1
ASCC 1275	1,845,495	39.1	1,974	1,666	15	55	234 (11.85)	106	95.5	100.0
APC151	1,839,134	39.1	1,982	1,687	18	67	206 (10.39)	106	95.5	100.0
ND03	1,831,949	39.0	1,968	1,692	15	57	200 (10.16)	105	94.6	99.1
B59671	1,821,173	39.1	1,925	1,567	18	67	269 (13.97)	106	95.5	100.0
EPS	1,812,305	39.0	1,937	1,608	18	67	240 (12.39)	106	95.5	100.0
LMG 18311	1,796,846	39.1	1,925	1,621	18	67	215 (11.17)	105	94.6	99.1
CNRZ1066	1,796,226	39.1	1,936	1,638	18	67	209 (10.80)	106	95.5	100.0
CS8	1,791,656	39.0	1,924	1,641	15	57	207 (10.76)	106	95.5	100.0
S9	1,787,436	39.1	1,922	1,630	18	67	203 (10.56)	106	95.5	100.0
ACA-DC 2	1,731,838	39.2	1,847	1,555	15	56	217 (11.75)	106	95.5	100.0

Cluster A strains start with strain JIM 8232 and end with strain ND03. Cluster B strains start with strain B59671 and end with strain ACA-DC 2. [1]Out of 111 essential genes. [2]Genome completeness as calculated by MiGA webserver considering 111 essential genes. [3]Corrected genome completeness considering 106 essential genes after the omission of glyS, proS, pheT, nahD, rpoC1 missing from all S. thermophilus genomes from the list of essential genes used by MiGA webserver.

Figure 1A). According to BPGA analysis, the *b* value of 0.14 in the power-law regression model is indicative of an open pan genome for *S. thermophilus* that is probably going to be closed soon (**Figure 1B**). This may also be supported by the fact that within the total of unique genes identified in *S. thermophilus* strains, 71% belong to three strains, namely KLDS 3.1003 (n = 41), JIM 8232 (n = 67), and NCTC12958^T (n = 204), while strains APC151, ASCC 1275, CNRZ1066, CS8, MN-BM-A01, MN-BM-A02, ND03, ND07, and S9 have no unique genes (**Supplementary Tables S1, S2**). BPGA analysis also revealed the number of exclusively absent genes per strain (**Supplementary Table S1**). Core, accessory and unique genes were further classified into COG categories, as implemented within the BPGA pipeline (**Supplementary Figure S2**). The analysis revealed that approximately 90% of the core, 60% of the accessory and 40% of the unique genes were assigned to various COG categories, with the rest having no prediction. We then excluded the poorly characterized categories R and S from further analysis. The majority of core genes encode proteins involved primarily in housekeeping and metabolic processes. The three most abundant COG categories were J (translation, ribosomal structure, and biogenesis, 12.7%), E (amino acid transport and metabolism, 11.8%), and L (replication, recombination, and repair, 7.3%). In the case of the accessory and unique genes the categories with the highest percentages included categories E, L, and K (transcription) and L, K, and V (defense mechanisms), respectively. In general, accessory and unique genes encoded among others transposases, Cas proteins, R-M systems, glycosyltranferases, polysaccharide biosynthesis proteins, amino acid biosynthesis proteins, proteolytic enzymes, stress related proteins, as well as transporters which may contribute to strain-specific technological traits (please see below).

The phylogenetic relationship among the *S. thermophilus* strains was determined based on the core genome of the strains and revealed two main clusters containing 15 (APC151, ASCC 1275, DGCC 7710, GABA, JIM 8232, KLDS 3.1003, KLDS SM, LMD-9, MN-BM-A01, MN-BM-A02, MN-ZLW-002, ND03, ND07, SMQ-301, and ST3; cluster A) and seven (ACA-DC 2, B59671, CNRZ1066, CS8, EPS, LMG 18311, and S9; cluster B) strains, respectively, while strain NCTC12958^T was placed separately (**Figure 2**). Moreover, the ANI phylogenetic tree had practically an identical topology to that of the phylogenetic tree (**Figure 3**). There was only one exception with strain KLDS 3.1003 being placed in cluster B. A more detailed inspection of the potential differences between strains in the two clusters revealed that cluster A strains had larger genomes beyond 1.83 Mbp, while those in cluster B had smaller genomes (**Table 2** and **Supplementary Figure S3**). This difference was found to

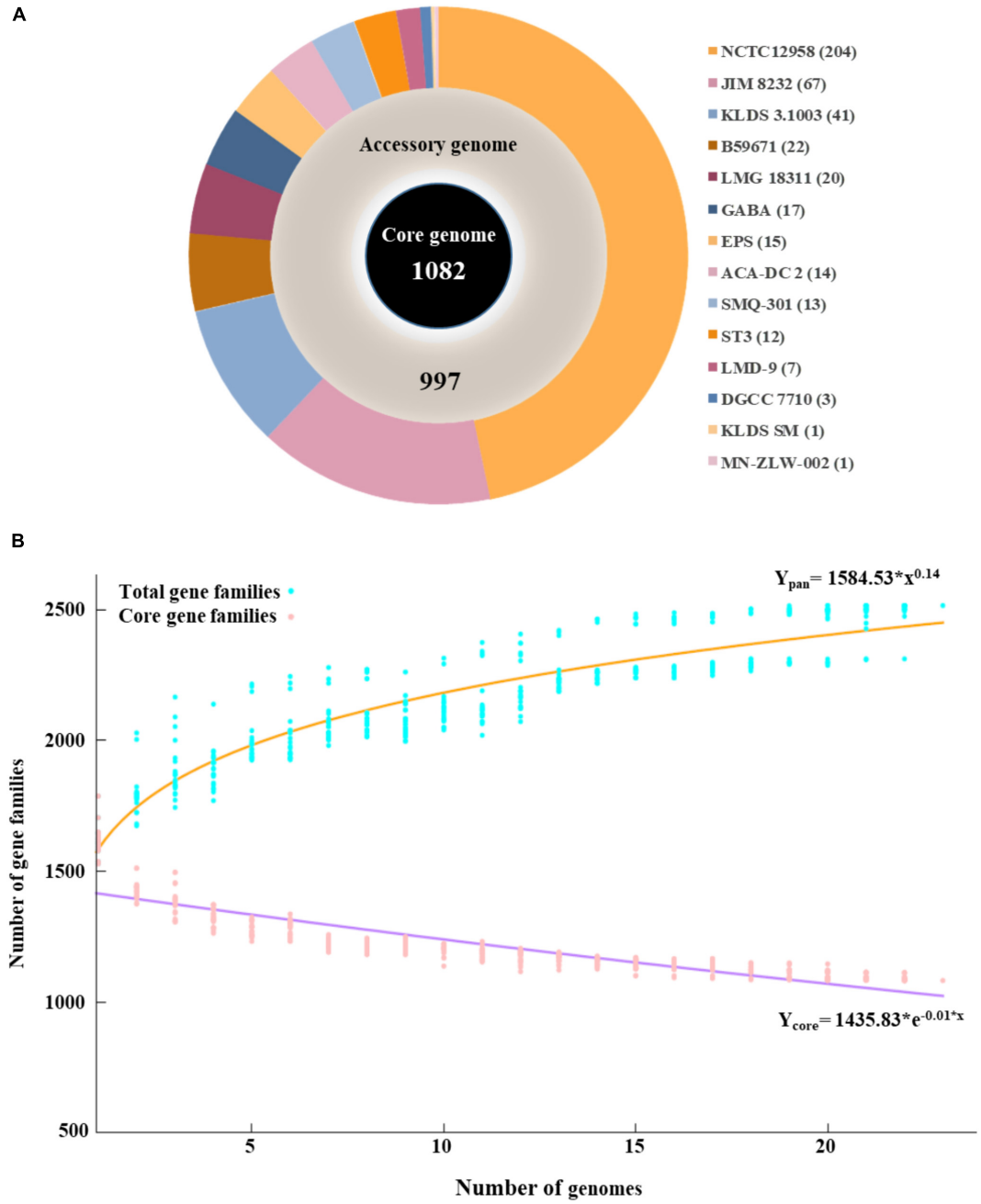


FIGURE 1 | Graphical presentation of core genome (inner black circle), accessory genome (middle gray circle) and unique genes (outer multicolored circle) of the 23 *S. thermophilus* strains. The number of unique genes for each strain is presented in the parentheses **(A)**. Pan and core genome plots of *S. thermophilus* strains analyzed **(B)**.

be statistically significant ($p < 0.05$) suggesting that strains in the two clusters may have been separated by distinct gene gain and/or gene loss events. Within these two main clusters, subgroups of *S. thermophilus* strains could also be identified during both phylogenetic and ANI analysis. These subgroups include strains ASCC 1275, DGCC 7710, KLDS SM, MN-BM-A02, and ND07 (subgroup AI), MN-BM-A01 and MN-ZLW-002 (subgroup AII), LMD-9 and SMQ-301 (subgroup AIII), APC151 and ND03 (subgroup AIV), and finally CNRZ1066, CS8, EPS, and S9 (subgroup BI) (**Figures 2**, **3**). As already mentioned, core genome phylogeny was also previously performed in a dataset of 25 *S. thermophilus* strains employing genomes sequenced to a variable degree of completeness (Delorme et al., 2017). In this study 1,311 core proteins were reported. Of note, an earliest study was performed based on three *S. thermophilus* genome sequences reporting 1,487 core genes (Lefébure and Stanhope, 2007). Our

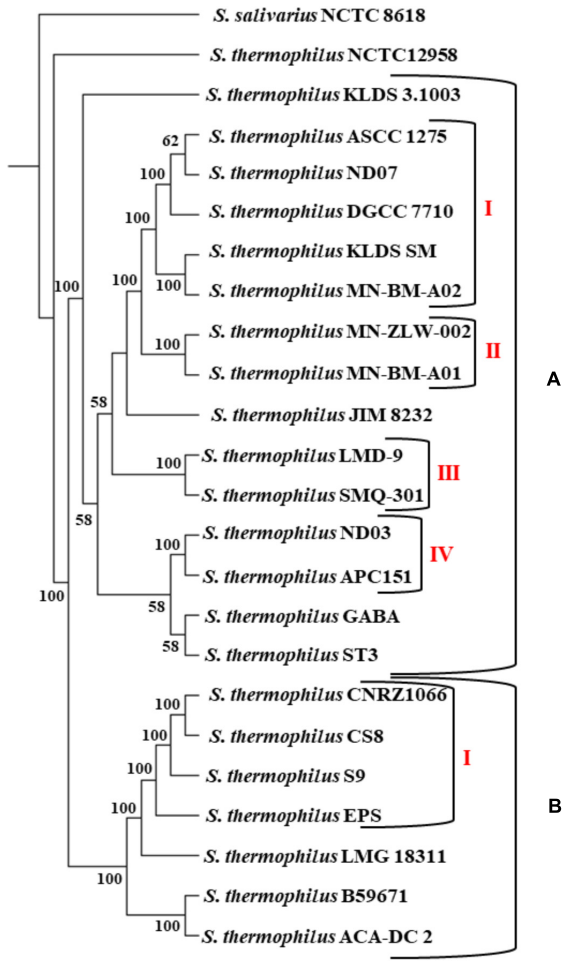


FIGURE 2 | Core genome phylogenetic tree of the 23 *S. thermophilus* strains. Strains were grouped in two clusters **(A,B)**. Subgroups within the clusters are also highlighted (AI, AII, AIII, AIV, and BI). *S. salivarius* NCTC 8618 was used as an outgroup. For branches with less than 50% bootstrap support the bootstrap values are not shown.

core genome was estimated to consist of 1,082 core proteins. This may suggest a more stringent selection of core proteins during our analysis. Despite the fact that several different strains were analyzed in our study and the study by Delorme et al. (2017), phylogenetic clustering of strains exhibited similarities supporting more or less the distinction we propose between cluster A and B strains and the subgroups observed within them. Differences in the topology of the two phylogenetic trees can be attributed to the different dataset of genomes analyzed as well as the different methods employed to construct the trees. The fact that we concentrated our analysis solely on strains with complete genome sequences presents an important advantage, since we were able to support clustering of strains based on the comparative genomic analysis of additional genomic traits as follows. Completeness of genome sequence is of utmost importance when the presence/absence of specific loci or their exact organization are the main factors for strain diversification.

The subgroups mentioned above appeared at high ANI values (>99.9%) which may suggest relatively subtle genomic differences. Such differences may indicate that strains of the same subgroup may be very similar but may deviate from the strict definition of clones. However, clonal relationships may be masked among strains due to aberrations in genome assembly that may come into play at such high ANI values (Burall et al., 2016). To avoid this pitfall, we investigated the quality of the assemblies of all *S. thermophilus* genomes analyzed in this study using the MiGA webserver (**Table 2**). Our analysis indicated that from the list of the 111 essential genes used to access genome completeness by MiGA, five (i.e., *glyS*, *proS*, *pheT*, *nhaD*, and *rpoC1*) were systematically missing from all *S. thermophilus* genomes. This observation suggested that they do not belong to the gene pool of the species, which is also supported by data presented previously for essential genes in Firmicutes (Albertsen et al., 2013). We thus corrected the completeness score of the genomes by calculating a total of 106 essential genes. Fifteen genomes received 100% genome completeness. Five genomes missed only *secE*, two missed *secE* plus an additional gene (*rpiX* or *uvrb*) and one missed only *ychF* receiving scores above 98.1%. The presence/absence frequency of *secE* may indicate that it is an accessory gene for *S. thermophilus*. In all cases the completeness scores of *S. thermophilus* genomes suggest perfect or nearly perfect assemblies. This is also corroborated by the quality scores for the genome assemblies that were all found "excellent" by MiGA webserver.

Hierarchical clustering of the COG frequency heat map generated for all *S. thermophilus* strains also supported the existence of the clusters and subgroups mentioned above, with minor alterations (**Figure 4**). Strains GABA and B59671 were placed in opposite clusters, while strains of the BI subgroup were associated more loosely (i.e., not forming a distinct subgroup). The most abundant category in all strains was E, followed by J and L. The prevalence of the E category may support adaptation of *S. thermophilus* to milk and the necessity of the organism to use amino acids from the environment.

The presence/absence heat map of the accessory genes of *S. thermophilus* strains supported once again the existence of clusters A and B (**Figure 5A**). The analysis allowed the identification of genes, which may contribute to the grouping of the strains. As shown in the horizontal axis of the heat map, genes within clusters 4 and 6 are characteristic of clusters B and A, respectively. Moreover, genes of clusters 1, 2, 3, 5, and 4 seem to be present in specific subgroups, namely AII, AIV, AIII, AI, and BI, respectively. Further analysis of the accessory proteins, specifically of those involved in metabolic processes, revealed that cluster A strains (including NCTC12958^{T}) carry the entire set of genes responsible for the biosynthesis of histidine that are basically absent from cluster B (**Figure 5B**). Based on these findings it is plausible to state that strains of *S. thermophilus* exhibit lineage-type relationships.

Lactose and Galactose Metabolism

Streptococcus thermophilus ferments preferentially lactose over glucose (Geertsma et al., 2005). Lactose is the main carbohydrate of milk and therefore constitutes the primary carbon and

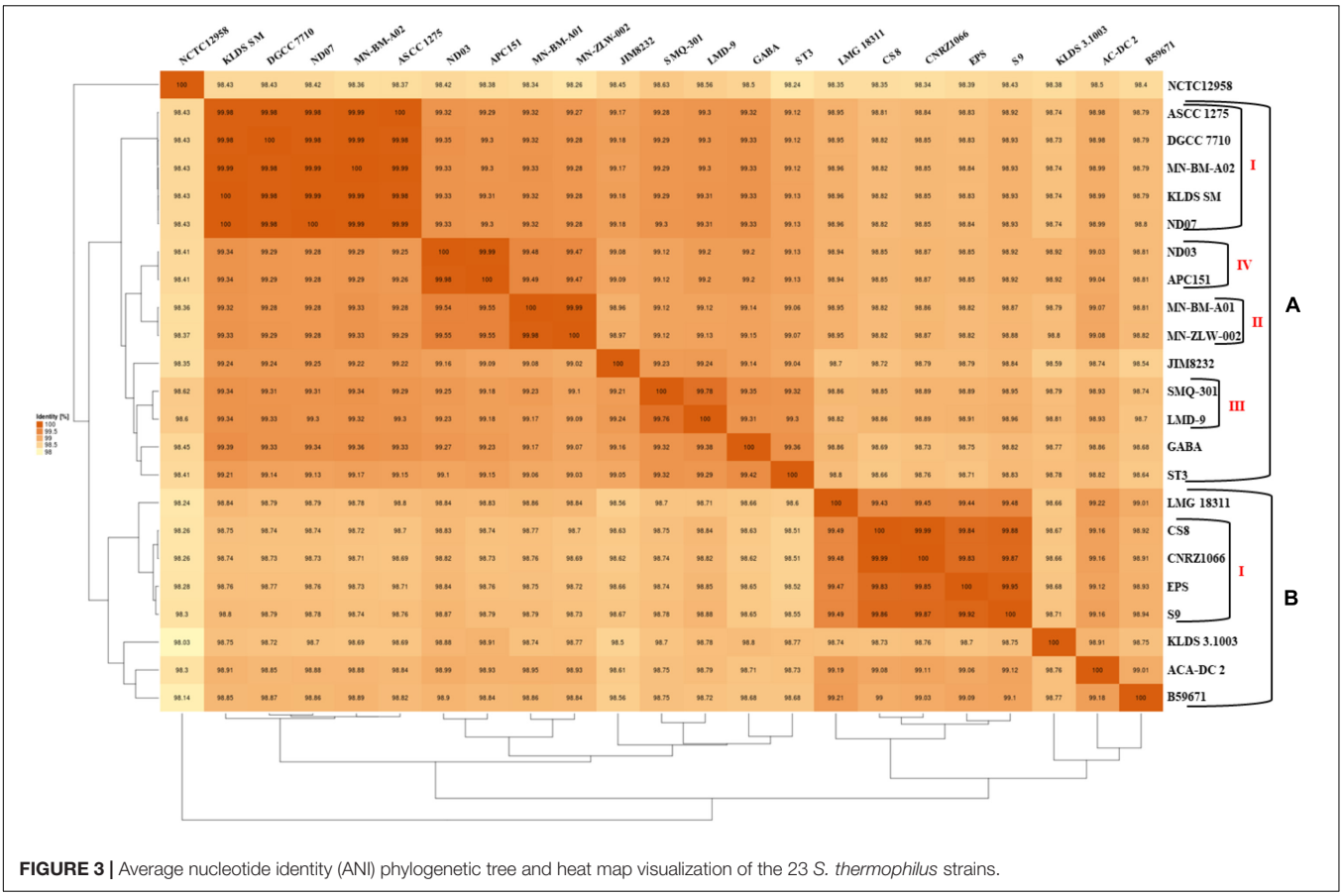


FIGURE 3 | Average nucleotide identity (ANI) phylogenetic tree and heat map visualization of the 23 *S. thermophilus* strains.

energy source for *S. thermophilus*, due to the adaptation of the microorganism to this particular niche (Bolotin et al., 2004; Hols et al., 2005; Goh et al., 2011). The genes implicated in the fermentation of lactose and galactose are organized in two adjacent operons (*galRKTEM-lacSZ*) (Vaughan et al., 2001). We found the complete locus in all *S. thermophilus* strains analyzed, with the exception of three strains in which *lacS* (strains B59671 and KLDS 3.1003) or *galR* (strain NCTC12958^{T}) are putative pseudogenes (**Supplementary Table S3**). The importance of these inactivations needs to be experimentally investigated, but the high degree of conservation of the *gal-lac* gene clusters among the different *S. thermophilus* strains, both at sequence and organization levels, reveals its importance in the catabolism of lactose in milk. Apart from *galE* coding for the enzyme UDP-glucose 4-epimerase that is located in the Leloir gene cluster, a second or even a third distal *galE* gene was identified in certain strains (**Supplementary Table S3**). It has been demonstrated that the activity of this enzyme is positively correlated with the biosynthesis of precursors for EPS production in EPS producing Gal^{-} *S. thermophilus* strains (Degeest and De Vuyst, 2000). Furthermore, the galactose moiety generated by the hydrolysis of lactose is translocated outside the cell via the dedicated antiporter LacS, which is implicated in the uptake of lactose in exchange to galactose (Vaughan et al., 2003). The majority of *S. thermophilus* strains are unable to metabolize both free and intracellularly produced galactose, probably either due to insufficient activities of *galK* and *galM* genes or due to mutations in the *galR-galK* promoter region, which may interfere with the expression levels of the respective enzymes (De Vin et al., 2005; Vaillancourt et al., 2008; Anbukkarasi et al., 2014; Sørensen et al., 2016). Recently, Xiong et al. (2019b) demonstrated that the Gal^{+} phenotype of *S. thermophilus* depends upon the expression of the *gal* operon, which can be widely affected by a single point mutation at the -9 box in the *galK* promoter. Since the accumulation of galactose in the medium by *S. thermophilus* may be important from a technological or nutritional perspective (Giaretta et al., 2018), we examined the presence of the mutation at the -9 box in the *galK* promoter in the strains analyzed. Accordingly, only B59671, CS8, EPS, and NCTC12958^{T} seem to be able to catabolize galactose, as they own the relevant G to A mutation in the position -9 of the -10 box related Gal^{+} phenotype (data not shown). However, experimental verification is required to validate this prediction.

Biosynthesis of EPS

One of the key technological properties of *S. thermophilus* is the production of EPS, which has been related to desirable textural properties and reduced syneresis in fermented dairy products (Lluis-Arroyo et al., 2014; Han et al., 2016). In a recent study, the EPS clusters of several strains were compared suggesting variations in the gene content of these loci (Cui et al., 2017). Our analysis revealed the presence of EPS gene clusters in all *S. thermophilus* strains examined. The size of

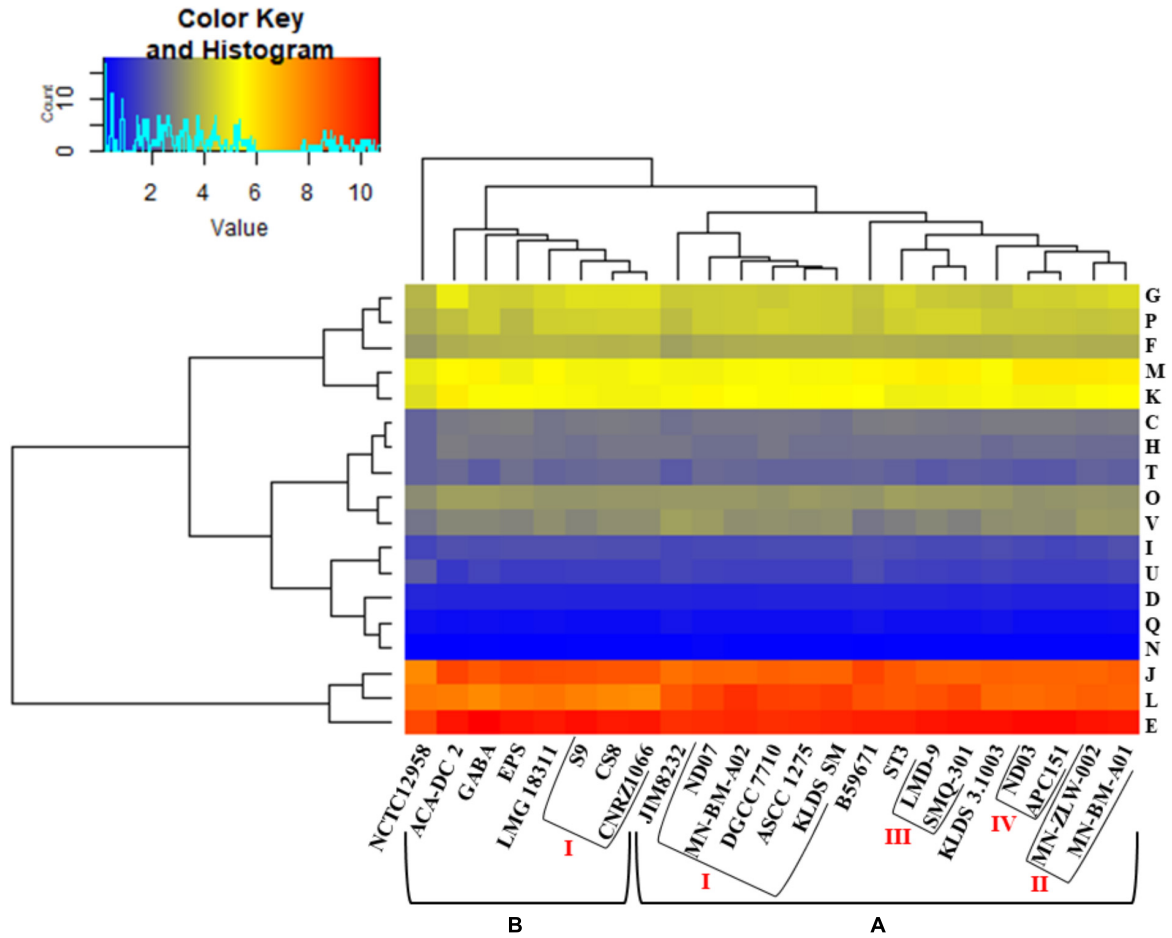


FIGURE 4 | Clusters of orthologous groups (COG) frequency heat map based on a two-dimensional hierarchical clustering. The horizontal axis corresponds to the percentage frequency of proteins involved in the respective COG functional categories: Information storage and processing: translation, ribosomal structure, and biogenesis (J), transcription (K), replication, recombination, and repair (L); cellular processes and signaling: cell cycle control, cell division, chromosome partitioning (D), cell wall/membrane/envelope biogenesis (M), cell motility (N), post-translational modification, protein turnover, chaperones (O), signal transduction mechanisms (T), intracellular trafficking, secretion, and vesicular transport (U), defense mechanisms (V); metabolism: energy production and conversion (C), amino acid transport and metabolism (E), nucleotide transport and metabolism (F), carbohydrate transport and metabolism (G), coenzyme transport and metabolism (H), lipid transport and metabolism (I), inorganic ion transport and metabolism (P), secondary metabolites biosynthesis, transport and catabolism (Q). The vertical axis shows the 23 *S. thermophilus* strains. Strains were grouped in two clusters **(A,B)**. Subgroups within the clusters are also highlighted (AI, AII, AIII, AIV, and BI). Categories R and S, concerning poorly characterized proteins, were not included in the analysis.

the clusters ranged between 18,661 and 35,973 bp and the % GC content (34.3–36.4%) was found to be lower than the % GC content calculated for the complete genomes of all strains (**Supplementary Table S4**). All clusters are flanked by a purine-nucleoside phosphorylase (*deo*D) and a transporter protein as their boundaries (**Figure 6**). The alignment of the EPS loci showed that they are highly conserved at the 5′ and the 3′ ends and their differences are located mainly in the middle of the clusters. At the 5′ end, genes *epsA*, *epsB*, *epsC*, and *epsD* were found in all EPS gene clusters and their role has been associated with the regulation of *eps* genes and chain elongation of the EPS molecules (Cui et al., 2017). The adjacent *epsE* gene coding a galactosyl-1-phosphate transferase was found in five out of 23 EPS gene clusters (strains CNRZ1066, CS8, EPS, S9, and SMQ-301). In the rest EPS clusters, *epsE* seems to encode a glycosyl-1-phosphate transferase. These enzymes initiate the assembly of the EPS repeating components through the transfer of phosphorylated sugars to the undecaprenyl-phosphate lipid carrier on the cytoplasmic side of the bacterial membrane (Broadbent et al., 2003; Wu et al., 2014). The sugar is transferred to the outer side of the membrane and this translocation process is probably facilitated by a flippase protein (Manat et al., 2014). All cluster A strains, including strain NCTC12958^{T}, carried one flippase coding gene with the exception of strain ST3 which carried two. In contrast, all strains from cluster B seem to lack the respective gene with the exception of strain B59671.

The genes downstream *epsE* encode proteins with various functions related to EPS biosynthesis. Among them, glycosyltransferases are involved in the consecutive transport of nucleotide sugar moieties to the lipid carrier. Both the number

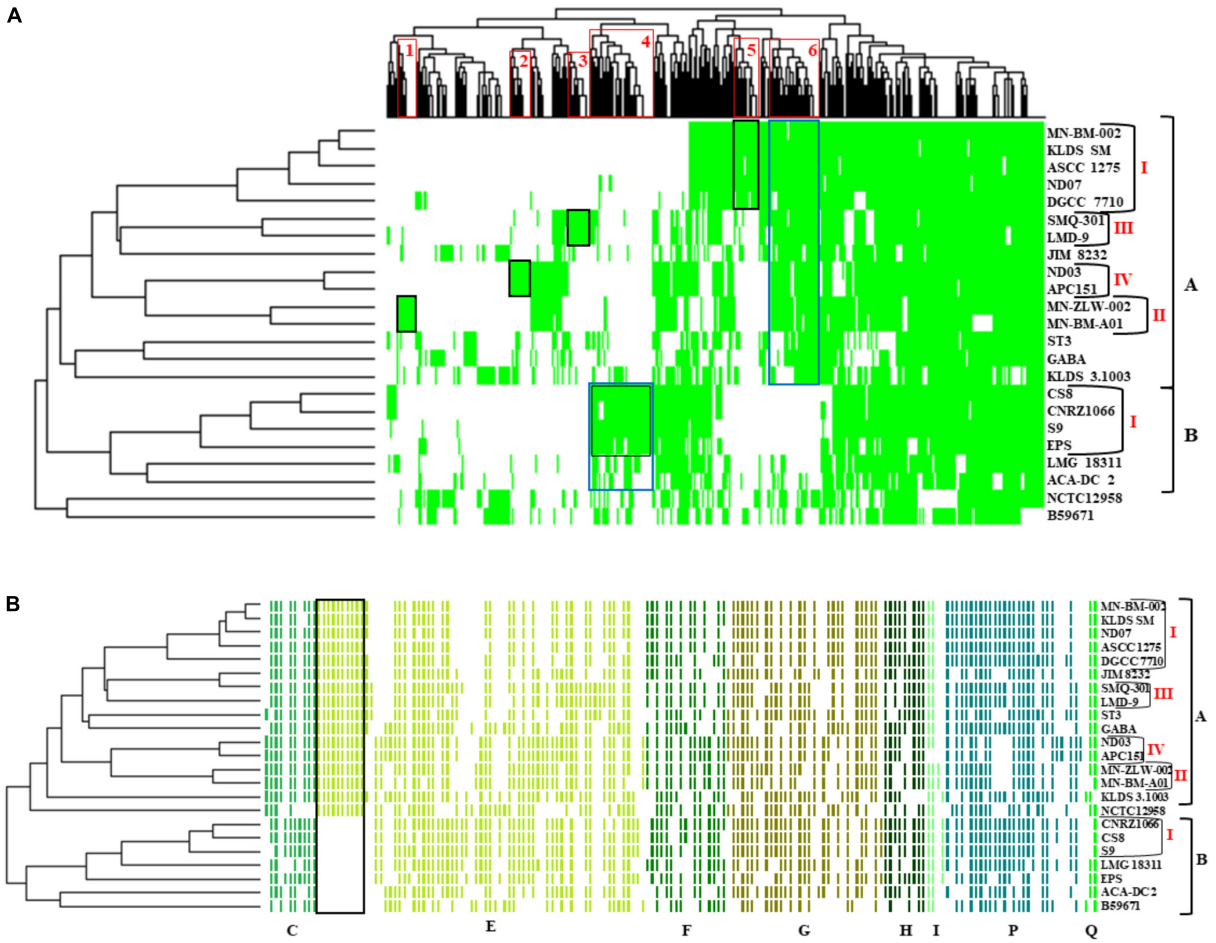


FIGURE 5 | Presence/absence heat map and two-dimensional hierarchical clustering of the 23 *S. thermophilus* strains accessory genome **(A)**. Colored areas represent the presence of genes in the respective *S. thermophilus* strains, while white areas indicate the absence of genes. Accessory genes clustered according to their presence in specific subgroups or clusters of strains (gene clusters 1 to 6) are highlighted with red frames on the *x* axis. The presence of a specific gene cluster in cluster A or B as well as specific subgroups of strains is highlighted with blue or black frames, respectively. Presence/absence heat map and hierarchical clustering of *S. thermophilus* strains based on accessory genes with clusters of orthologous groups (COG) assignment involved in metabolism (categories C, E, F, G, H, I, P, and Q) **(B)**. Colored areas represent the presence of genes in the respective *S. thermophilus* strains, while white areas indicate the absence of genes. Genes implicated in the biosynthesis of histidine are highlighted with a black frame. Of note, panel **(B)** is a composite figure of an excel generated table manually colored, while clustering was exported from RStudio. This was necessary to achieve clustering of genes grouped based on COG categories.

and the type of the respective genes in the EPS clusters are variable and may influence the composition of the produced EPS (Cui et al., 2017). The current analysis revealed the presence of transferases commonly encountered in *S. thermophilus* EPS clusters, such as glucosyltransferases, galactosyltransferases, and rhamnosyltransferases. A UDP-galactopyranose mutase involved in the synthesis of UDP-galactofuranose was identified in half of the EPS clusters. Interestingly, only strains JIM 8232, GABA, and NCTC12958^T were found to carry a gene encoding a putative galactofuranosyl-transferase. Finally, genes implicated in the polymerization and translocation of the EPS repeating units have been also identified in all EPS clusters, as reported previously for *S. thermophilus* strains (Goh et al., 2011; Wu et al., 2014; Cui et al., 2017; Evivie et al., 2017).

Based on synteny, the EPS gene clusters can be categorized practically in distinct groups, supporting AI, AII, AIV, and BI subgroups. EPS clusters of strains KLDS 3.1003 and ST3 were highly similar to subgroup AI. Similarities in EPS clusters were also observed beyond lineages, as in the case of strains GABA and NCTC12958^T. Certain EPS gene clusters, namely those of strains ACA-DC 2, LMD-9, LMG 18311, and B59671, presented higher structural variability due to the presence of many unique genes, which are coding mostly hypothetical proteins and glycosyltransferases. These observations are in accordance with previous findings for strains LMD-9 and LMG 18311 (Goh et al., 2011). Of note, three recent studies have been performed to highlight molecular mechanisms of EPS production in strains ASCC 1275 and KLDS SM (Li et al., 2018; Padmanabhan et al., 2018; Wu and Shah, 2018), while a fourth study suggests a protective role of purified EPS isolated from strain MN-BM-A01 against colitis in mice (Chen et al., 2019). The lineage like-patterns we observed among EPS gene clusters could potentially

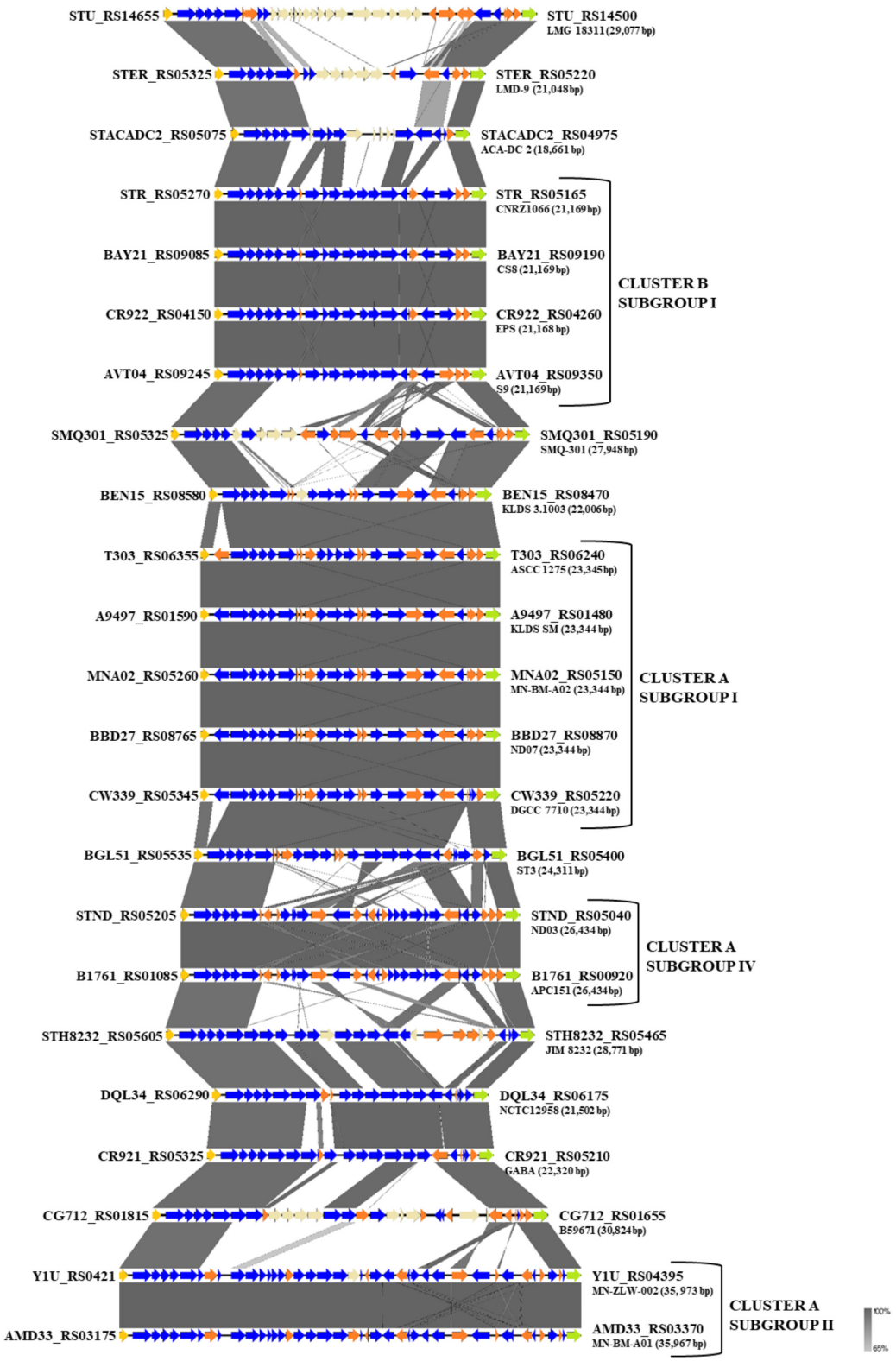


FIGURE 6 | Multiple sequence alignment of the exopolysaccharide (EPS) gene clusters of the 23 *S. thermophilus* strains after BLASTN analysis. Gray shading represents the % identity among the nucleotide sequences according to the color gradient presented at the lower right corner of the figure. Protein coding genes are highlighted in dark blue, putative pseudogenes in orange, the *deoD* in yellow, the transporter gene in green and the unique genes for each *S. thermophilus* strain in beize. Clusters and subgroups of strains are highlighted.

be useful for extrapolating findings from one strain to another. In all cases understanding of the EPS biosynthesis in *S. thermophilus* may allow a better selection of strains or even their engineering for improved dairy and probiotic products (Xiong et al., 2019a).

Proteolytic System

The proteolytic system of LAB has been extensively investigated. A number of studies have revealed the diversity of its components, i.e., cell-wall bound proteinases, peptide and amino acid transporters and peptidases, among various LAB species (Savijoki et al., 2006; Liu et al., 2010). In the present study, the proteolytic system of *S. thermophilus* strains was examined on the basis of the scheme published by Liu et al. (2010) and the recent work of Tian et al. (2018). The results acquired from the TransportDB database were also employed.

Due to the limited availability of free amino acids and peptides in milk, the degradation of caseins is essential for growth. In *S. thermophilus*, the cell-wall associated proteinase PrtS is implicated in the initiation of the proteolytic cascade (Hols et al., 2005; Goh et al., 2011; Tian et al., 2018). *prtS* is present in almost half of the strains examined. The analysis showed that the respective gene is present (intact or truncated) solely in cluster A strains with the exception of strains APC151, KLDS 3.1003, and ND03 (**Supplementary Table S5A**). As it has been previously reported, PrtS presents 95% identity to the PrtS protein of *Streptococcus suis* and the distribution of *prtS* in *S. thermophilus* strains is infrequent in historical collections compared to industrial ones, indicating acquisition by lateral transfer in the species population (Delorme et al., 2010). PrtS has been related to the rapid growth of *S. thermophilus* in milk as a mono-culture and therefore in the rapid acidification of milk, which is a desirable technological trait. However, the sole presence of *prtS* is not sufficient for the rapid milk acidification by *S. thermophilus*. Milk acidification seems to be a complex phenotypic trait, which involves the overexpression of several genes (Galia et al., 2016). Furthermore, it was demonstrated that *S. thermophilus* strains, irrespective of the $prtS^{+/-}$ status, may present cell-associated extracellular peptidase activities. These activities, albeit weaker than that of PrtS, could probably provide amino acids essential for *S. thermophilus* growth (Hafeez et al., 2015). The extracellular presence of PepX aminopeptidase in *S. thermophilus* was recently suggested (Hafeez et al., 2019). Nevertheless, it has been supported that only $prtS^{-}$ *S. thermophilus* strains can perform protocooperation with *L. bulgaricus* (Settachaimongkon et al., 2014).

Several peptide and amino acid transporters of various families have been predicted in all *S. thermophilus* strains (**Supplementary Table S5B**). The majority of these transporters belong to the ATP-binding cassette (ABC) superfamily and include one oligopeptide Opp ABC transporter, one branched-chain amino acid ABC transporter, one glutamine ABC transporter, four amino acid ABC transporters, one spermidine/putrescine ABC transporter and one methionine ABC transporter. In a number of instances, the gene clusters of these transporters may contain putative pseudogenes and thus may be not functional. It has been previously reported that strain LMD-9 carries a second Opp ABC transporter, which is homologous to that of *Bifidobacterium* species (Goh et al., 2011). This transporter is also present in strains SMQ-301 and ST3. Strains B59671, GABA, and NCTC12958^{T} have one extra amino acid ABC transporter, which displays high identity (90%) with the respective one of *S. salivarius* (data not shown). Furthermore, all strains carry four amino acid permeases of the amino acid-polyamine-organocation (APC) family. Additionally, strains ACA-DC 2, APC151, B59671, GABA, KLDS 3.1003, and ND03 carry a glutamate/GABA antiporter (*gadC*) (**Supplementary Table S5B**). The latter gene along with glutamate decarboxylase gene (*gadB*) are responsible for gamma-aminobutyric acid (GABA) production. It was recently demonstrated that strain APC151 is a high-yield GABA producer (Linares et al., 2016). In strain KLDS 3.1003 a unique histidine/histamine antiporter has been also identified (*hdcP*) (**Supplementary Table S5B**). The respective gene is located adjacently to a unique histidine decarboxylase gene (*hdcA*) and along with *hdcB* form the *hdc* cluster, probably acquired by HGT (please see below) which has been previously described in two other strains of *S. thermophilus* (Calles-Enríquez et al., 2010). From a physiological point of view, this gene cluster is probably implicated in cell protection under acidic conditions (De Angelis and Gobetti, 2011). The use of histamine-producing *S. thermophilus* strains should be avoided in dairy manufacture, since it has been demonstrated that $hdcA^{+}$ *S. thermophilus* used as starter in cheese production was associated with the accumulation of histamine in the final product (Gardini et al., 2012). One di-tripeptide transporter is present in all strains. A branched-chain amino acid permease and an amino efflux protein are also present in all strains, but for B59671 and ST3, respectively. The transport of the branched-chain amino acids leucine, isoleucine, and valine, as well as alanine, serine/threonine and glutamate/aspartate is probably facilitated by four symporters, three of them being present in all strains and only one in six strains (**Supplementary Table S5B**). In addition, a number of incomplete ABC transporters has been also predicted in all the strains analyzed (data not shown).

Besides PrtS, 12 highly conserved cytoplasmic peptidases have been identified in all strains, namely *pepA*, *pepC*, *pepF*, *pepM*, *pepN*, *pepO*, *pepP*, *pepQ*, *pepS*, *pepT*, *pepV*, and *pepX* (**Supplementary Table S5A**). Moreover, a number of peptidases, which have been identified in several LAB species, are missing from all *S. thermophilus* strains (Liu et al., 2010). More specifically, pyrrolidone-carboxylate peptidase (*pcp*) and proline peptidases *pepI*, *pepR*, and *pepL* are absent. Cysteine aminopeptidase (*pepE/pepG*) presents 40% identity with aminopeptidase C in all *S. thermophilus* strains, while a putative dipeptidase *pepD* is present but truncated in 14 *S. thermophilus* strains. It should be mentioned that the universal distribution of the majority of genes encoding proteins of the proteolytic system of *S. thermophilus* supports the essential role of the system.

Amino Acids Biosynthesis

The *in silico* analysis of amino acid biosynthetic pathways has been addressed in *S. thermophilus* (Hols et al., 2005). Experimental data for the species have been acquired for the biosynthesis of proline, branched-chain amino acids, glutamine and aspartate (Limauro et al., 1996; Garault et al., 2000; Monnet

et al., 2005; Arioli et al., 2007). Furthermore, Pastink et al. (2009) studied the amino acid metabolism and amino acid dependency of strain LMG 18311 through amino acid omission experiments, concluding that the minimal amino acid auxotrophy for the strain involves histidine and one of the sulfur-containing amino acids (methionine or cysteine). In some *S. thermophilus* strains amino acid requirements for growth involve at least four amino acids (Glu, Cys, His, and Met; Letort and Juillard, 2001). It seems that amino acid auxotrophy may be a strain dependent trait.

Most amino acid biosynthetic pathways are highly conserved in the 23 *S. thermophilus* strains (**Supplementary Figure S4** and **Supplementary Table S6**). Analysis of *S. thermophilus* protein coding sequences, based on KEGG orthology assignments and Hols et al. (2005), revealed that the majority of the amino acid biosynthetic pathways are present in all strains examined. Complete biosynthetic pathways in all *S. thermophilus* strains were predicted for threonine, cysteine, glycine, proline, glutamine, asparagine, phenylalanine, alanine, aspartate, and glutamate. Current annotations of all *S. thermophilus* strains in Refseq with prokaryotic genome annotation pipeline (PGAP) do not seem to support biosynthesis of lysine due to the absence of *dapE*, *dapH*, and *dapF* (**Supplementary Table S6**). An incomplete Dap-pathway was also reported for strains LMG 18311 (Hols et al., 2005) and LMD-9 (Goh et al., 2011). However, experimental evidence suggests biosynthesis of lysine in strains LMG 18311 (Pastink et al., 2009) and MN-ZLW-002 (Qiao et al., 2018) presumably through a complete Dap-pathway. We found that this discrepancy may be an artifact of annotation with the PGAP tool. Older *S. thermophilus* GenBank files, annotated with tools other than PGAP included a locus with three genes, the second of which is identified as a (truncated) *dapE* (data not shown). In contrast, in the same locus, PGAP predicts a single gene corresponding to a putative M20 peptidase pseudogene (e.g., locus_tag Y1U_RS01580 in strain MN-ZLW-002). We also tested other annotation tools, like rapid annotation using subsystem technology (RAST; Aziz et al., 2008) and FGenesB (Solovyev and Salamov, 2011) that also supported a three-gene architecture in the same locus, suggesting that further investigation is required to resolve this matter.

The most striking difference in the biosynthesis of amino acids among *S. thermophilus* strains examined concerns histidine. Hols et al. (2005) reported absence of this gene cluster in strains CNR1066 and LMG 18311 but its presence in strain LMD-9. As mentioned above, the respective pathway is complete in strains of cluster A and strain NCTC12958^T, while strains of cluster B carry only one related gene, namely *hisK* (**Supplementary Table S6** and **Figure 5B**). Furthermore, several amino acid biosynthetic pathways seem to be incomplete in a number of strains. Analysis revealed that in strain B59671 several genes involved in amino acid biosynthesis are putative pseudogenes or absent compared to the other strains. In this strain glutamate, serine, methionine and tyrosine biosynthetic pathways may be non-functional. Concerning the rest of the strains analyzed, incomplete biosynthetic pathways have been identified for methionine in NCTC12958^T and ST3, arginine in MN-BM-A01, branched-chain amino acids in JIM 8232 and tryptophan in EPS (**Supplementary Table S6**).

In some cases, differences among genes involved in specific biosynthetic steps during amino acid biosynthesis have been also identified. In tryptophan biosynthesis, two adjacent genes, namely *aroG1* and *aroG2* (Hols et al., 2005), encoding 70% identical proteins, have been identified in all strains except for strains ST3, CNRZ1066, and CS8. The first strain carries only *aroG1*, while the last two only *aroG2*. These genes are involved in the first step of chorismate synthesis, an intermediate product during tryptophan biosynthesis. Concerning the biosynthesis of branched-chain amino acids, in all *S. thermophilus* genomes two *ilvD* genes have been identified; one belongs to the *ilvDBNC* operon, while the second is located remotely from the *ilvDBNC* locus and its functionality is yet to be studied (Hols et al., 2005). The *ilvD* within the operon is a putative pseudogene in most strains and it seems to be functional only in KLDS 3.1003, LMD-9, NCTC12958^T, and SMQ-301. These observations need further experimental investigation.

Urea Metabolism

Streptococcus thermophilus is perhaps the sole species among the dairy LAB with the ability to hydrolyze urea, a phenotypic trait, which affects adversely the milk acidification rate (Pernoud et al., 2004; Iyer et al., 2010). The urease gene cluster is highly conserved in all *S. thermophilus* strains analyzed and comprises 11 genes in the form of an operon of 8.2 kbp size (**Supplementary Table S7**). It includes the acid-activated *ureI* gene, the structural genes *ureABC*, the accessory genes *ureEFGD* and the genes encoding the cobalt/nickel uptake system *ureMQO* (or *cbiMQO*) (Mora et al., 2004; Iyer et al., 2010). The *ureI* gene is located upstream the structural genes and is coding a pH-dependent urea channel, which is probably activated for compensating the increase of the extracellular acidity. The *ureABC* genes are coding the three structural subunits of the enzyme, with *ureC* coding the large subunit and the remaining two genes coding the two smaller subunits (Ninova-Nikolova and Urshev, 2013). The auxiliary genes *ureEFGD* encode metallochaperones involved in nickel metallocenter biosynthesis and the delivery of nickel ions to the active site of the urease. More specifically, the urease apoenzyme forms a complex with the UreD, UreF, and UreG proteins, which is activated by the addition of nickel, bicarbonate and the metallochaperone UreE (Sujoy and Aparna, 2013). The *ureMQO* system is probably responsible for the translocation of nickel ions into the bacterial cell as indicated by functional analysis of the homologous genes in *S. salivarius* (Chen and Burne, 2003).

The physiological role of *S. thermophilus* urease has not been thoroughly evaluated. Although it is considered a response mechanism to acid stress, it has been demonstrated that urease is produced at low levels also at neutral pH (Mora et al., 2005). The ureolytic activity of *S. thermophilus* is probably related not only to the biosynthesis of essential amino acids, e.g., glutamine, but to the overall nitrogen metabolism of the species, with the expression of the *ure* operon depending on aspartate, glutamate, glutamine, and NH_3 concentrations (Monnet et al., 2005; Arioli et al., 2007). However, the rather uncommon urease-negative phenotype has been also reported for *S. thermophilus* strains,

indicating that urease activity may not hold a vital role in milk fermentation (Mora et al., 2002). Recently, spontaneous urease-deficient mutants of *S. thermophilus* were isolated from *S. thermophilus* populations deriving from industrial yogurt starters. The stability of the mutated phenotype was confirmed, providing promising results regarding the potential use of urease-deficient strains as starters in dairy fermentations (Ninova-Nikolova and Urshev, 2013). However, in a recent study employing urease deficient mutants it was suggested that urease activity is important for yogurt acidification and that its absence inhibits fermentation acceleration during protocooperation with *L. bulgaricus* (Yamauchi et al., 2019).

CRISPR-Cas Systems

The CRISPR-Cas systems are defense mechanisms widely distributed in prokaryotes, providing acquired immunity against foreign genetic elements like viruses and plasmids (Horvath and Barrangou, 2010). This immunity mechanism has been extensively studied in *S. thermophilus*, providing information concerning the environmental adaptability and the anti-phage activity of this microorganism (Sapranauskas et al., 2011; Louis et al., 2017; Hao et al., 2018). In addition, in certain studies spacers within CRISPR arrays in *S. thermophilus* were employed for assessing diversity among strains of the species (Horvath et al., 2008; Delorme et al., 2017). As mentioned above, Delorme et al. (2017) reported that MLST and whole genome based phylogeny differed from those inferred by CRISPR analysis. Here we revisit clustering of *S. thermophilus* strains based on CRISPR analysis in the context of complete genome sequences that allowed us further validation of the diversity scheme we propose in this study.

As reported previously (Horvath and Barrangou, 2010), up to four distinct CRISPR-Cas loci, i.e., CRISPR1, CRISPR2, CRISPR3, and CRISPR4 were identified in our *S. thermophilus* strains (**Supplementary Tables S8, S9A** and **Figure 7**). CRISPR1 and CRISPR3 both belong to Class 2/subtype II-A CRISPR-Cas systems, while CRISPR2 and CRISPR4 belong to Class 1/subtype III-A and Class 1/subtype I-E CRISPR-Cas systems, respectively (Horvath et al., 2008; Makarova et al., 2015; Hao et al., 2018). Furthermore, one putative orphan CRISPR array structure was predicted by CRISPRFinder in strains JIM 8232 and LMG 18311, characterized by the absence of adjacent Cas proteins. The direct repeats (DRs) of this array in JIM 8232 were identical to the DRs of CRISPR3 in other strains, suggesting that it must have owned the relevant Cas proteins originally and subsequently lost them. In contrast, the DRs of LMG 18311 in the orphan array did not match any other DRs.

CRISPR1 was found in 22 out of the 23 *S. thermophilus* strains analyzed here, with ACA-DC 2 carrying no CRISPR array despite retaining CRISPR-related genes (Alexandraki et al., 2017). CRISPR1 array size ranged between 760 and 2,805 bp. This size variability is associated with the number of spacers (11–42) in the arrays of the different strains. This is the largest CRISPR array within the *S. thermophilus* strains analyzed with the exception of strain ST3 (**Figure 7**) and it has been reported to be ubiquitous in *S. thermophilus* strains (Horvath et al., 2008). In strains B59671 and KLDS 3.1003 the gene coding the Cas9 protein is a putative pseudogene, indicating that the respective CRISPR-Cas systems might have been inactivated. Strains ASCC 1275, APC151, DGCC 7710, GABA, KLDS 3.1003, KLDS SM, LMD-9, MN-BM-A02, MN-ZLW-002, ND03, ND07, NCTC12958^T, SMQ-301, and ST3 also carry CRISPR3. This CRISPR contains 8 to 26 spacers and in most cases is shorter than CRISPR1 (**Figure 7**). A higher activity for CRISPR1 in comparison to CRISPR3 has been experimentally validated (Horvath et al., 2008). In the case of CRISPR3, *cas9* is a putative pseudogene in strain MN-BM-A01, indicating that the specific system may have been also inactivated. It should be emphasized that CRISPR1 is detected in both cluster A and B strains, while CRISPR3 is present only in cluster A (apart from strain JIM 8232) and it is totally absent from cluster B strains. Based on analysis of the LMD-9 genome sequence, Horvath et al. (2008) proposed that the entire CRISPR3-Cas system may have been deleted or inserted in *S. thermophilus* strains through a recombination event between a repeat present in the terminal repeat of CRISPR3 and a repeat close to *serB* which flanks the system from one side.

CRISPR2 was found in strains ASCC 1275, DGCC 7710, GABA, KLDS 3.1003, KLDS SM, JIM 8232, LMD-9, LMG 18311, MN-BM-A02, ND07, SMQ-301, and ST3. Thus, CRISPR2 was present only in cluster A strains, apart from strain LMG 18311 which belongs to cluster B. Among the Cas proteins of CRISPR2, *cas1* is a putative pseudogene in strains KLDS 3.1003 and LMG 18311, while *cas10* is a putative pseudogene in strains LMD-9 and SMQ-301. Furthermore, the respective CRISPR-Cas systems of strains GABA and ST3 carry only three CRISPR-associated genes (*cas1*, *cas2*, and *csm6*) which indicates that they are incomplete. All these CRISPR2 systems carried a CRISPR array. However, additional "possible" CRISPR2 systems were predicted by CRISPRFinder owning an incomplete set of Cas proteins followed by a single spacer within two DRs (**Supplementary Table S9A**). Our findings suggest inactivation and/or degeneration of CRISPR2 in several strains. Horvath et al. (2008) reported that CRISPR2 may indeed be inactivated in certain strains, however Tamulaitis et al. (2014) were able to demonstrate its activity in at least another strain. CRISPR4 was identified in strains ASCC 1275, B59671, DGCC 7710, KLDS SM, MN-BM-A02, and ND07. Genes *cse1* and *cas2* in strains ASCC 1275 and B59671, respectively, are putative pseudogenes. Interestingly, the CRISPR4 was basically found in subgroup AI strains. Further subgrouping could be supported not only through the presence/absence of CRISPR-Cas systems, but also through the distribution of different spacers, as discussed below.

A total of 997 spacers were found in the confirmed CRISPR-Cas systems of the 22 *S. thermophilus* strains with 93% being assigned in CRISPR1 and CRISPR3. Analysis of the respective sequences revealed that 258 are unique among 11 strains, namely NCTC12958^T, JIM 8232, GABA, KLDS 3.1003, ST3, B59671, LMG 18311, LMD-9, SMQ-301, S9, and EPS, while 253 appeared more than once in the CRISPR arrays. As shown previously, CRISPR arrays may be employed for accessing strain diversity within *S. thermophilus* (Horvath et al., 2008; Delorme et al., 2017). Indeed, looking into the architecture of the CRISPR arrays we could identify once more patterns that are not shared by all *S. thermophilus* strains, but they are specific to the grouping of strains we have already described. For example, CRISPR1

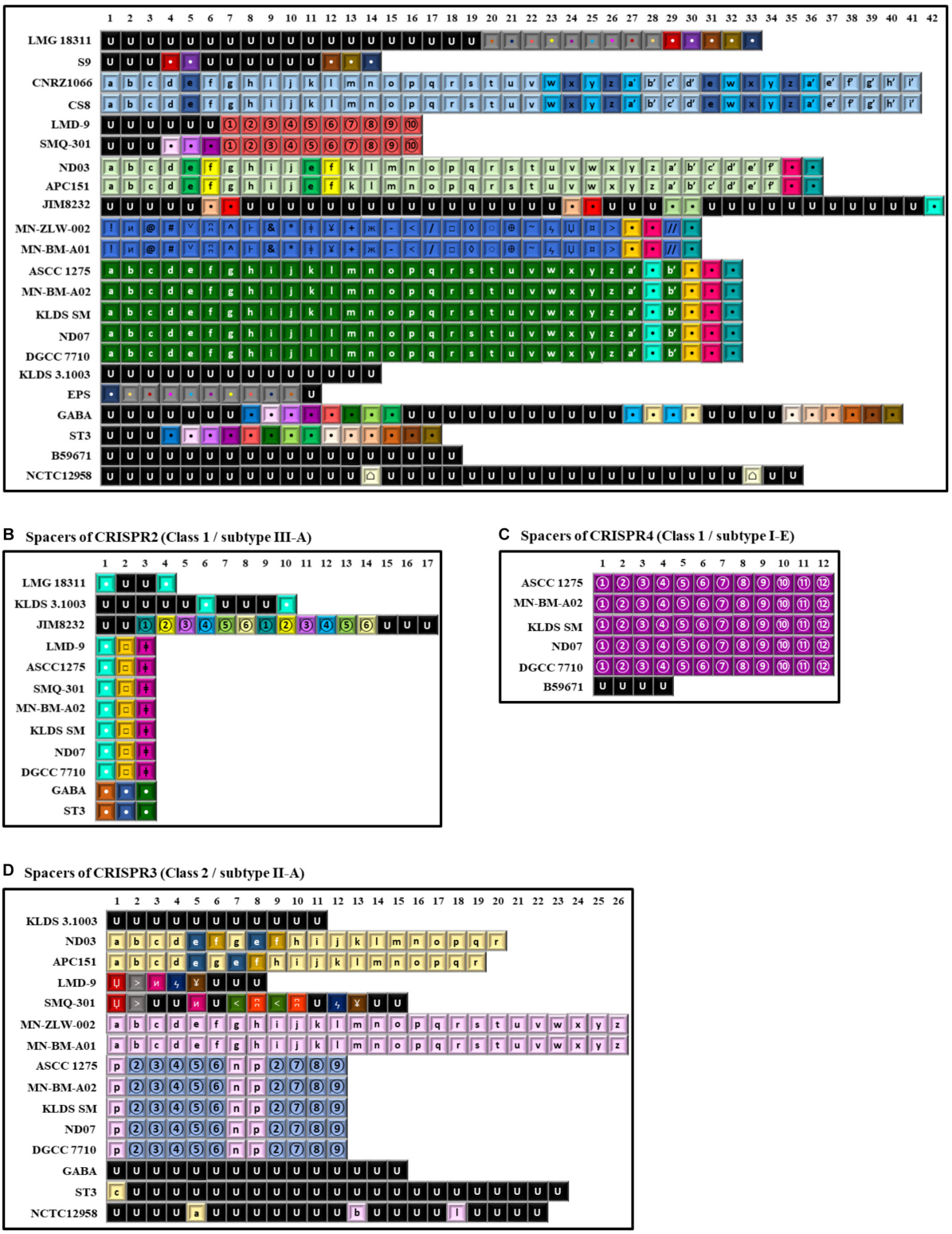


FIGURE 7 | Spacer sequences alignment of the various clustered regularly interspaced short palindromic repeats-CRISPR associated (CRISPR-Cas) system types found in the 22 *S. thermophilus* strains. In the alignments only the spacer sequences have been used. In each type of CRISPR-Cas system each spacer is represented by the combination of a character and a font color. The spacers represented in black font with the letter U correspond to unique spacers. Spacers represented by the same combination of a character and a font color correspond to identical spacers. Spacers of CRISPR1 **(A)**, CRISPR2 **(B)**, CRISPR4 **(C)**, and CRISPR3 **(D)**.

supports subgroups AI, AII, AIII, and AIV. Subgroup BI is partially supported, since only strains CNRZ1066 and CS8 share the same CRISPR array. CRISPR3 supports subgroups AI, AII, AIII, and AIV. CRISPR4 has a unique pattern of spacers for subgroup AI. As mentioned above CRISPR2 is present only in cluster A strains, apart from strain LMG 18311 which belongs to cluster B, but the spacer pattern in the arrays could not distinguish any subgroup (**Figure 7**). Most spacers were unique for each subgroup and were present in a specific order in the array. This observation suggests that this part of the array was present in the common ancestor of these subgroups of strains. However, in certain instances, a specific spacer could be found common between two seemingly unrelated arrays belonging to different subgroups of strains. Most probably such spacers were acquired by the common ancestor of each subgroup due to exposure to the same exogenous DNA that resulted in the acquisition of the same part of sequence into the specific CRISPR array. Evidently, these spacers were identified only in arrays of the same class and subtype CRISPR-Cas systems. Similar analysis of spacers to infer evolutionary relationships among *S. thermophilus* strains have been reported previously (Horvath et al., 2008). However, when looking solely to the architecture of the CRISPR array it is very difficult to distinguish between clones or complexes of very similar strains that are not actual clones.

BLASTN analysis of the spacers showed that 317 sequences matched several different *S. thermophilus* bacteriophages (**Supplementary Table S9B**). Almost half of the spacers analyzed could be related to phages 7201, Sfi19, Sfi21, DT1, and Sfi11. This finding may indicate a high frequency of exposure of *S. thermophilus* to the specific phages. Finally, six spacers were highly identical to *Lactococcus* phages, while 12 spacers were highly identical to plasmids of *Enterococcus faecium*, *S. suis*, *Streptococcus pyogenes*, *S. pneumoniae*, *Lactobacillus salivarius* and *Lactococcus lactis*. These findings indicate that *S. thermophilus* has been found in the same environment with these bacteria. Furthermore, it could be hypothesized that at least some potential HGT events of plasmid donation toward *S. thermophilus* were aborted through the activity of CRISPR-Cas systems. Overall our findings are in agreement with previous results (Bolotin et al., 2005; Horvath et al., 2008).

R-M Systems and Prophages

Another immunity mechanism employed by the prokaryotes against foreign DNA are the R-M systems. All *S. thermophilus* strains analyzed carry several R-M systems, classified into four types (Roberts et al., 2005, 2015; **Supplementary Table S10** and **Supplementary Figure S5**). The majority of strains carry one complete type I R-M system with strains EPS, NCTC12958^T, GABA, and KLDS 3.1003 carrying two. No type I R-M system was predicted for strain B59671, while in strains MN-ZLW-002, MN-BM-A01, and ST3 the predicted type I R-M system was incomplete due to the absence or inactivation of one or more of the necessary genes. This was the case for additional predicted type I systems in several strains. Certain *S. thermophilus* strains carry at least one type II system with strains LMD-9, MN-BM-A01, ND03, and APC151 owning three such systems. Unlike type I R-M systems, most type II systems seem to be complete and potentially active. One type III system is present in strains ACA-DC 2, CNRZ1066, CS8, EPS, S9, LMG 18311, NCTC12958^T, and GABA. Finally, a type IV system has been predicted in almost half of the strains analyzed, which contains only a restriction enzyme that recognizes and cuts modified DNA.

A more detailed investigation revealed that type II and type III R-M systems are absent or inactivated from *S. thermophilus* strains of subgroup AI. For this reason, we wanted to examine whether the presence/absence pattern of R-M systems in *S. thermophilus* is lineage specific. As demonstrated in **Supplementary Figure S5** in several instances R-M systems are distributed on the chromosome in a manner that is characteristic for a potential lineage. This is particularly obvious for cluster A and more specifically for subgroups I, II, III, and IV. The R-M systems of strains in cluster B presented some similarities within the same subgroup, but they were more variable.

Despite the presence of the aforementioned defense mechanisms, complete prophages were also predicted for strains APC151, ND03, and NCTC12958^T, while in the rest of the examined genomes only remnants of prophages have been identified (data not shown). In strains APC151 and ND03 the same prophage was predicted located within the EPS cluster of each strain. In strain NCTC12958^T the intact prophage was previously described as phage 20617 by Arioli et al. (2018). Interestingly, the authors of this study demonstrated that the lysogenic strain NCTC12958^T (DSM 20617^T) exhibited higher adhesion to solid surfaces and heat resistance compared to the phage-cured derivative strain, suggesting some competitive advantage due to the stable association of the phage and the host.

Genomic Islands

Genomic islands acquired through HGT can provide adaptive and technological traits to the host microorganism (Juhas et al., 2009). *In silico* prediction of HGT in *S. thermophilus* has been previously reported (Hols et al., 2005; Liu et al., 2009; Eng et al., 2011). In this study, the GIs predicted by IslandViewer 4 in *S. thermophilus* ranged from 5 to 23 per strain, with sizes between 3.5 and 58 kbp and variable GC content from 26.1 to 45.2% (**Supplementary Table S11**). A total of 253 GIs were predicted, 31 of which were unique in 11 strains. The rest of the GIs have been identified in at least two *S. thermophilus* strains, either complete or partial. Of note, the genome array of ribosomal proteins was predicted as part of a GI in a number of strains. This is a false positive result, since it has been reported that the nucleotide composition of these arrays differentiates significantly from the rest protein coding genes (Hols et al., 2005; Fernandez-Gomez et al., 2012). Thus, these GIs (14 in total) were excluded from further analysis (**Supplementary Tables S11, S12A**). Several GIs were found to be present in both clusters A and B strains, while others were present either in cluster A or B strains. The first type of GIs was most probably acquired earlier than the second, i.e., before clusters A and B were separated. In accordance with what has been reported above for other genomic features, certain subgroups of strains display a unique distribution pattern of specific GIs that can support subgroups AI–AIV and BI (**Supplementary Figure S6**).

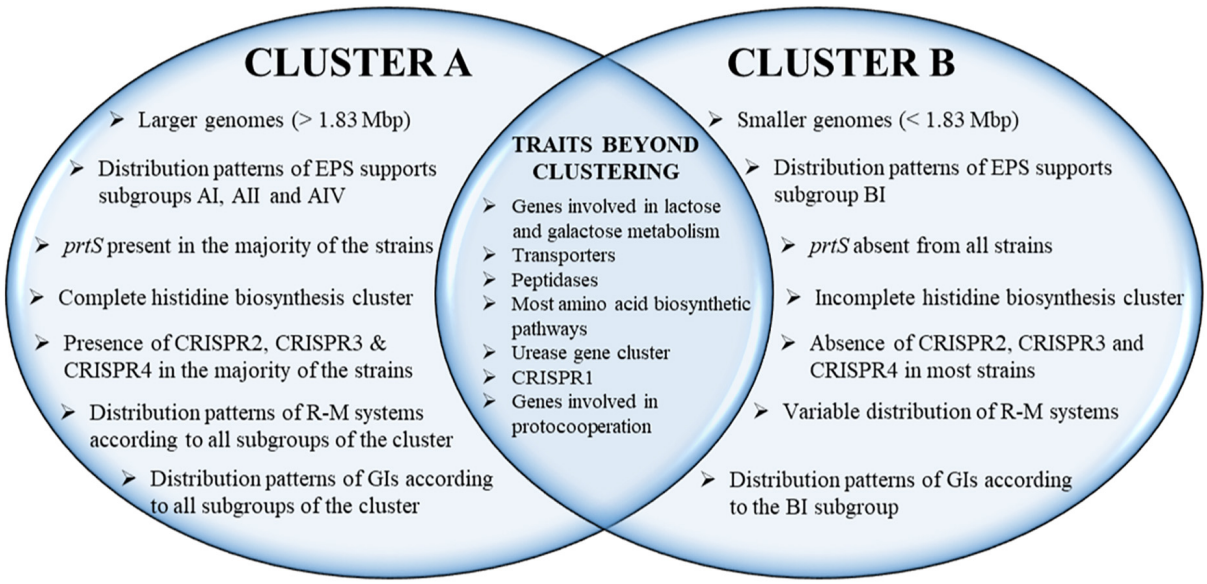


FIGURE 8 | Schematic representation of structural and functional genomic traits that support the distinction of *S. thermophilus* strains in clusters A and B. The genomic features included in the common area of the Venn diagram are either present to all strains or they exhibit a presence/absence pattern beyond clusters A and B.

BLASTN analysis of the predicted GIs could not always reveal a potential donor. Nevertheless, a number of GIs could be traced back to specific microorganisms (coverage >70%, identity >90%; **Supplementary Table S12B**). The majority of species acting as potential donors belongs to the *Streptococcus* genus but also to other LAB like *L. lactis*, *Lactobacillus casei*, and *Leuconostoc gelidum*. In these last three cases GIs present high identity to plasmids carried by these organisms. In detail, specific GIs in subgroups AI, AII, and one GI in strain NCTC12958^T present high identity to plasmids pLd7/p229C of *L. lactis* subsp. *lactis* (Kelleher et al., 2017; Van Mastrigt et al., 2018), pBD-II/pLC2W of *L. casei* (Ai et al., 2011; Chen et al., 2011; Song et al., 2018) and plasmid 1 of *L. gelidum* subsp. *gasicomitatum* (Andreevskaya et al., 2016), respectively. It is interesting to highlight that strains of *S. thermophilus* seem to have also interacted with members of the *Streptococcus bovis*/*Streptococcus equinus* complex (SBSEC), namely *S. macedonicus*, *S. infantarius* subsp. *infantarius*, *Streptococcus gallolyticus*, and *S. equinus*. Members of the complex are established members of the gastrointestinal tract (GIT) of ruminants, while certain species like *S. macedonicus* and *S. infantarius* are increasingly associated with fermented foods, especially of dairy origin (Jans et al., 2013a,b; Papadimitriou et al., 2014, 2015a).

A detailed investigation of the annotated features of *S. thermophilus* GIs revealed that they could be involved in EPS biosynthesis in accordance with previous findings reported for strains CNRZ1066, LMD-9, and LMG 18311 (Liu et al., 2009). CRISPR-Cas and complete R-M systems have been also identified in GIs. This would include CRIPR3 and CRISPR4 and type I and III R-M systems. In addition, the 38.5 kbp GI 9 contains most part of the intact prophage in strain NCTC12958^T (**Supplementary Table S12A**). Our analysis supports the presence of bacteriocin coding genes in the GIs of a number of strains. However, Hols et al. (2005) suggested that the activity of these antimicrobial peptides may not be always guaranteed due to the absence of genes coding for transport or immunity proteins or other differences. For example, the locus of a class II bacteriocin-like peptide (*blp*) was experimentally studied in strains CNRZ1066, LMG 18311, and LMD-9 and it was concluded that it is only functional in the last strain (Hols et al., 2005). In strain B59671, GI 5 carries genes of the *blp* gene cluster involved in the production of the bacteriocin thermophilin 110 (Renye et al., 2017). Finally, in GI 6 of strain GABA we found a locus containing several genes coding for leader peptides (including mutacin IV, BlpU, and bovicin 255), but transport or immunity proteins seem to be inactive or absent (**Supplementary Table S12A**). Moreover, several genes involved in amino acid transport have been found in the predicted GIs of *S. thermophilus* strains. Some of these include a glutamate:GABA antiporter in strains APC151, GABA, and ND03, a dicarboxylate/amino acid:cation symporter in strains APC151, KLDS 3.1003, MN-BM-A01, MN-ZLW-002, ND03, and ST3 and a complete amino acid ABC transporter in strains CS8, EPS, KLDS 3.1003, and S9. The *hdc* cluster of strain KLDS 3.1003 was also identified in a GI and BLASTN analysis revealed possible HGT from a satellite phage. Furthermore, GI 7 of strain JIM 8232 corresponds to the biosynthetic gene cluster of histidine. As already mentioned, this region is also present in all cluster A *S. thermophilus* strains (plus strain NCTC12958^T) but for unknown reasons it was assigned as a GI only in JIM 8232. BLASTN analysis revealed that this region presents high identity to the SBSEC member *S. equinus* (92%) supporting its potential acquisition by HGT in *S. thermophilus* chromosome. In addition, genes involved in fatty acid biosynthesis were identified in GIs of strains

APC151, GABA, MN-BM-A01, MN-ZLW-002, and ND03, while stress response genes, e.g., coding for cold-shock proteins were also identified in a number of strains, including ASCC 1275, CNRZ1066, KLDS 3.1003, LMG 18311, ND03, and ST3. Finally, the gene cluster *cbs-cblB-cysE* involved in the metabolism of sulfur-containing amino acids has been previously suggested to have been transmitted by HGT from *L. bulgaricus* or *Lactobacillus helveticus* to *S. thermophilus* (Liu et al., 2009). Current analysis revealed that the respective cluster was predicted as part of a bigger GI in 17 *S. thermophilus* strains. More specifically, this GI along with the three genes were identified in strains APC151, GABA, KLDS 3.1003, LMD-9, LMG 18311, MN-BM-A01, MN-ZLW-002, ND03, and SMQ-301, while in strains ACA-DC 2, ASCC 1275, CNRZ1066, CS8, MN-BM-A02, ND07, S9, and ST3 the *cysE* is a putative pseudogene (**Supplementary Table S12A**).

It should be mentioned that Selle et al. (2015) identified four expendable GIs in the genome of strain LMD-9 with variable distribution in other sequenced strains. IslandViewer 4 did not predict GIs 1 and 2 reported in that study, while it detected GIs overlapping or included in GIs 3 and 4. These differences can be explained by the *in silico* methods employed to detect GIs. Selle et al. (2015) employed a strategy combining the location of potentially essential open reading frames (ORFs) and highly similar insertion sequences (ISs) which is distinct from the strategies employed by the tools included in IslandViewer 4.

S. thermophilus Genes Implicated in Protocooperation With *L. bulgaricus*

The bacterial pair of *S. thermophilus* and *L. bulgaricus* is routinely employed in yogurt production. The mutually beneficial interaction between these bacteria in the yogurt ecosystem, known as protocooperation, is based on the exchange of metabolites and results in improved metabolic performance related to accelerated acidification, enhanced EPS production and abundance of aroma volatiles. Initially, *S. thermophilus* boosts the growth of *L. bulgaricus* by lowering the pH and providing formic, pyruvic and folic acid as well as carbon dioxide. Subsequently, *L. bulgaricus* stimulates *S. thermophilus* growth by producing peptides and free amino acids (Settachaimongkon et al., 2014). Transcriptome analysis of a mixed *S. thermophilus* and *L. bulgaricus* culture also supports that metabolites like formic and folic acid produced by *S. thermophilus* are utilized by *L. bulgaricus* as precursors in purine biosynthesis (Sieuwerts et al., 2010). *S. thermophilus* carries genes encoding pyruvate formate lyase (PFL) and pyruvate formate-lyase activating (PFLA) enzyme, while *L. bulgaricus* lacks these genes (Nishimura et al., 2013). Our analysis revealed the presence of both *pfl* and *pflA* in all *S. thermophilus* strains examined (**Supplementary Table S13**).

In addition, a number of studies have been performed concerning the role of PrtS produced by *S. thermophilus* during manufacture of dairy products, especially yogurt. For example, PrtS production may positively affect *S. thermophilus* growth in a pure culture, but it may be neutral in a mixture with *L. bulgaricus* strains producing the protease PrtB (Courtin et al., 2002). In a more recent study, it was demonstrated that only non-proteolytic *S. thermophilus* strains performed protocooperation with *L. bulgaricus* (Settachaimongkon et al., 2014). As already mentioned, the majority of cluster A strains carries *prtS*, while it is absent from all cluster B strains, indicating that the latter may be more appropriate for protocooperation. However, specific *S. thermophilus* strains carrying the *prtS* have been shown to exhibit weak or no PrtS activity (Galia et al., 2009; Cui et al., 2016). In our dataset in strains MN-BM-A01 and SMQ-301 *prtS* was found to be truncated, an observation that may support to a degree the findings by Galia et al. (2009). Furthermore, it was recently reported that $prtS^+$ strains may also present some technological advantages (Tian et al., 2018). We thus believe that more research is needed to establish the actual role of *prtS* regarding protocooperation.

The response of *S. thermophilus* to H_2O_2 produced by *L. bulgaricus* has also been studied. It appears that there is an inverse correlation between iron intake by *S. thermophilus* and H_2O_2 production by *L. bulgaricus*, and that *S. thermophilus* in the presence of H_2O_2 is regulating iron metabolism in order to diminish the production of harmful reactive oxygen species (ROS) (Herve-Jimenez et al., 2009; Sieuwerts et al., 2010). However, the results of two different studies are rather diverge. In one study, the expression patterns of *S. thermophilus* genes related to iron transport in the presence of *L. bulgaricus* were found to be upregulated (Sieuwerts et al., 2010), while in another study downregulated (Herve-Jimenez et al., 2009). Only *dpr* (peroxide resistance protein) and *fur* (ferric transport regulator protein) were found upregulated in both studies. *In silico* analysis of the 23 *S. thermophilus* strains revealed that *dpr* and *fur* belong to the core genome, while the iron ABC transporter is absent from strains JIM 8232, MN-ZLW-002, ND03, APC151, MN-BM-A01, and ST3 (**Supplementary Table S13**).

A novel protocooperation relationship between *S. thermophilus* and *L. bulgaricus* in yogurt fermentation concerns the bi-functional glutathione (GSH) synthetase gene of *S. thermophilus*, which produces GSH (Wang et al., 2016). The respective gene was found to be conserved in all 23 *S. thermophilus* strains analyzed (**Supplementary Table S13**). In a recent study, it was demonstrated that GSH produced by *S. thermophilus* provided protection to both *S. thermophilus* and *L. bulgaricus* cells toward acid stress. Additionally, the secreted GSH could enhance the growth of *L. bulgaricus* (Wang et al., 2016). Finally, genes related to EPS production were found to be upregulated in both microorganisms in a mixed culture when compared to monocultures, and thus they may play an important role in the texture of the final product (Sieuwerts et al., 2010). Given the heterogeneity observed in the EPS gene cluster of *S. thermophilus* strains, no mechanistic insight could be inferred.

CONCLUSION

Streptococcus thermophilus is a starter of great economic significance for the dairy industry contributing to the production of world-wide consumed dairy products like yogurt and cheeses. A number of studies have been published in an attempt to explore and interpret various features of the species biology related to its technological potential. This became more feasible

during the last two decades with the sequencing of genomes of *S. thermophilus* strains. In this study we analyzed 23 fully sequenced genomes of *S. thermophilus* in order to examine features of the species related to technological and evolutionary traits. Even from the beginning of our study, it became evident that strains of *S. thermophilus* present some variability considering the properties of the genomes (e.g., size, gene content, % of pseudogenes, rRNA and tRNA content). Core genome and ANI phylogenetic analysis revealed a specific pattern of clustering of strains (**Figure 8**). A main observation was that most strains could be separated in two major clusters. Cluster A was characterized by larger genomes, the presence of *prtS* in the majority of strains, the inclusion of a histidine biosynthesis gene cluster, as well as the presence of certain CRISPR-Cas system types and specific GIs. Strains in cluster B diversified from those in cluster A in all these aspects. These observations indicated the existence of at least two major lineages in *S. thermophilus* that appear at ANI values >98%. Further investigation suggested the presence of subgroups within the two clusters, i.e., subgroups AI–AIV and BI. The existence of these subgroups was also supported to a variable degree during COG analysis as well as the presence/absence pattern of specific loci and/or their organization, i.e., EPS clusters, CRISPR arrays, R-M systems and GIs. Clustering of *S. thermophilus* strains based on the spacers of CRISPR arrays has been performed before (Horvath et al., 2008; Delorme et al., 2017). Given the fact that CRISPR arrays can provide a retrospective view of the history of each strain based on the parasitic DNA it was exposed to, spacer sequences of the CRISPR1 which is present practically in all strains support the existence of evolutionary distinct lineages in *S. thermophilus*. Biodiversity within strains of *S. thermophilus* has been previously suggested using CRISPR array and/or MLST clustering (Horvath et al., 2008; Delorme et al., 2010, 2017; Yu et al., 2015). In our opinion, clustering of strains according to CRISPR array architecture or even MLST has important advantages (e.g., the ability to screen many strains), but these approaches may derive more easily to the characterization of potential clonal strains due to the use of limited genomic information. In contrast, whole genome phylogeny based on core genes should be more robust, while analysis of complete genome sequences may provide even more information concerning the discrimination of strains based on loci beyond core genome, like accessory genes or even unique genes. The subgroups we describe appeared at ANI values well above 99%, an observation that could indicate that they derive from clonal strains. A closer investigation of the data presented in this study suggests in some cases differences among strains of the same subgroup. For example, this becomes obvious when considering the exact sizes of the chromosome of the strains, the exact gene content (including accessory genes but also genes that are exclusively absent from a specific strain). In some instances, differences were observed in the EPS clusters, the distribution of R-M systems and GIs of strains within the same subgroup. Even though the differences among strains of the same subgroup may be rather subtle thus justifying the high ANI values at which their relatedness appears, they diversify strains beyond the strict definition of clones. Our analysis concerning the genome assemblies of the strains suggested a quality level that may not interfere with the grouping scheme we describe. Nonetheless, apart from the differences identified among the strains, our analysis also validated common features or features beyond the clustering pattern mentioned above (**Figure 8**). These would include characteristic traits for the adaptation of *S. thermophilus* to milk, like the conserved *gal-lac* and urease operons, the extended arsenal of peptidases and amino acid/peptide transporters in parallel to genes related to protocooperation. The high percentage of pseudogenes has been related to the reductive evolution of *S. thermophilus* during adaptation to rich in nutrients dairy niches (Bolotin et al., 2004; Hols et al., 2005; Goh et al., 2011). This trait was also apparent in all strains analyzed here. Interestingly, features related to milk adaptation seem to be also present in APC151. The strain does not diversify from the dairy strains, even though it was the only strain in our dataset that was isolated from a non-dairy environment, i.e., the fish intestine. This was also suggested previously (Linares et al., 2017). This relatively odd observation highlights the need to study strains found in environments different than milk and dairy products to fully apprehend the evolution of the species. Finally, the pan genome of the species is not closed yet, suggesting that sequencing of additional strains will be important. Certain new complete genomes have appeared in the databases since the initiation of our analysis (Proust et al., 2018; Renye et al., 2019), but more are required to further expand and validate any lineage-like patterns that may exist and could be related to the technological/probiotic repertoire of *S. thermophilus*.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

VA and MK performed genome analysis and participated in the writing of the manuscript. JB and BP performed genome analysis. KP conceived the project, performed genome analysis, and participated in the writing of the manuscript. ET conceived the project and participated in the writing of the manuscript. All authors read and approved the final manuscript.

FUNDING

The present work was co-financed by the European Social Fund and the National Resources EPEAEK and YPEPTH through the Thales project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02916/full#supplementary-material

REFERENCES

Ai, L., Chen, C., Zhou, F., Wang, L., Zhang, H., Chen, W., et al. (2011). Complete genome sequence of the probiotic strain *Lactobacillus casei* BD-II. *J. Bacteriol.* 193, 3160–3161. doi: 10.1128/JB.00421-11

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538. doi: 10.1038/nbt.2579

Alexandraki, V., Kazou, M., Blom, J., Pot, B., Tsakalidou, E., and Papadimitriou, K. (2017). The complete genome sequence of the yogurt isolate *Streptococcus thermophilus* ACA-DC 2. *Stand. Genomic Sci.* 12:18. doi: 10.1186/s40793-017-0227-5

Anbukkarasi, K., Nanda, D. K., Umamaheswari, T., Hemalatha, T., Singh, P., and Singh, R. (2014). Assessment of expression of Leloir pathway genes in wild-type galactose-fermenting *Streptococcus thermophilus* by real-time PCR. *Eur. Food Res. Technol.* 239, 895–903. doi: 10.1007/s00217-014-2286-9

Anbukkarasi, K., Umamaheswari, T., Hemalatha, T., Nanda, D. K., Singh, P., Rashmi, H. M., et al. (2013). Production of low browning Mozzarella cheese: screening and characterization of wild galactose fermenting *Streptococcus thermophilus* strains. *Int. J. Adv. Res.* 1, 83–96.

Andreevskaya, M., Hultman, J., Johansson, P., Laine, P., Paulin, L., Auvinen, P., et al. (2016). Complete genome sequence of *Leuconostoc gelidum* subsp. *gasicomitatum* KG16-1, isolated from vacuum-packaged vegetable sausages. *Stand. Genomic Sci.* 11:40. doi: 10.1186/s40793-016-0164-8

Arioli, S., Eraclio, G., Della Scala, G., Neri, E., Colombo, S., Scaloni, A., et al. (2018). Role of temperate bacteriophage ϕ20617 on *Streptococcus thermophilus* DSM 20617T autolysis and biology. *Front. Microbiol.* 9:2719. doi: 10.3389/fmicb.2018.02719

Arioli, S., Monnet, C., Guglielmetti, S., Parini, C., De Noni, I., Hogenboom, J., et al. (2007). Aspartate biosynthesis is essential for the growth of *Streptococcus thermophilus* in milk, and aspartate availability modulates the level of urease activity. *Appl. Environ. Microbiol.* 73, 5789–5796. doi: 10.1128/AEM.00533-07

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75

Bai, Y., Sun, E., Shi, Y., Jiang, Y., Chen, Y., Liu, S., et al. (2016). Complete genome sequence of *Streptococcus thermophilus* MN-BM-A01, a strain with high exopolysaccharides production. *J. Biotechnol.* 224, 45–46. doi: 10.1016/j.jbiotec.2016.03.003

Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing Group, Lau, B. Y., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343

Blom, J., Kreis, J., Spanig, S., Juhre, T., Bertelli, C., Ernst, C., et al. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* 44, W22–W28. doi: 10.1093/nar/gkw255

Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S. D., Kulakauskas, S., et al. (2004). Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat. Biotechnol.* 22, 1554–1558. doi: 10.1038/nbt1034

Bolotin, A., Quinquis, B., Sorokine, A., and Dusko Ehrlich, S. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–2561. doi: 10.1099/mic.0.28048-0

Broadbent, J. R., Mcmahon, D. J., Welker, D. L., Oberg, C. J., and Moineau, S. (2003). Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. *J. Dairy Sci.* 86, 407–423. doi: 10.3168/jds.S0022-0302(03)73619-4

Burall, L. S., Grim, C. J., Mammel, M. K., and Datta, A. R. (2016). Whole genome sequence analysis using JSpecies tool establishes clonal relationships between *Listeria monocytogenes* strains from epidemiologically unrelated listeriosis outbreaks. *PLoS One* 11:e0150797. doi: 10.1371/journal.pone.0150797

Calles-Enríquez, M., Eriksen, B. H., Andersen, P. S., Rattray, F. P., Johansen, A. H., Fernández, M., et al. (2010). Sequencing and transcriptional analysis of the *Streptococcus thermophilus* histamine biosynthesis gene cluster: factors that affect differential *hdcA* expression. *Appl. Environ. Microbiol.* 76, 6231–6238. doi: 10.1128/AEM.00827-10

Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* 6:24373. doi: 10.1038/srep24373

Chen, C., Ai, L., Zhou, F., Wang, L., Zhang, H., Chen, W., et al. (2011). Complete genome sequence of the probiotic bacterium *Lactobacillus casei* LC2W. *J. Bacteriol.* 193, 3419–3420. doi: 10.1128/JB.05017-11

Chen, Y., Zhang, M., and Ren, F. (2019). A role of exopolysaccharide produced by *Streptococcus thermophilus* in the intestinal inflammation and mucosal barrier in Caco-2 monolayer and dextran sulphate sodium-induced experimental murine colitis. *Molecules* 24:513. doi: 10.3390/molecules24030513

Chen, Y.-Y. M., and Burne, R. A. (2003). Identification and characterization of the nickel uptake system for urease biogenesis in *Streptococcus salivarius* 57.I. *J. Bacteriol.* 185, 6773–6779. doi: 10.1128/jb.185.23.6773-6779.2003

Courtin, P., Monnet, V., and Rul, F. (2002). Cell-wall proteinases PrtS and PrtB have a different role in *Streptococcus thermophilus*/*Lactobacillus bulgaricus* mixed cultures in milk. *Microbiology* 148, 3413–3421. doi: 10.1099/00221287-148-11-3413

Cui, Y., Jiang, X., Hao, M., Qu, X., and Hu, T. (2017). New advances in exopolysaccharides production of *Streptococcus thermophilus*. *Arch. Microbiol.* 199, 799–809. doi: 10.1007/s00203-017-1366-1

Cui, Y., Xu, T., Qu, X., Hu, T., Jiang, X., and Zhao, C. (2016). New insights into various production characteristics of *Streptococcus thermophilus* strains. *Int. J. Mol. Sci.* 17:E1701. doi: 10.3390/ijms17101701

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147

Darling, A. E., Miklós, I., and Ragan, M. A. (2008). Dynamics of genome rearrangement in bacterial populations. *PLoS Genet.* 4:e1000128. doi: 10.1371/journal.pgen.1000128

De Angelis, M., and Gobetti, M. (2011). "Stress responses of lactobacilli," in *Stress Responses of Lactic Acid Bacteria*, eds E. Tsakalidou, and K. Papadimitriou, (New York, NY: Springer Science+Business Media), 219–249. doi: 10.1007/978-0-387-92771-8_11

De Vin, F., Radstrom, P., Herman, L., and De Vuyst, L. (2005). Molecular and biochemical analysis of the galactose phenotype of dairy *Streptococcus thermophilus* strains reveals four different fermentation profiles. *Appl. Environ. Microbiol.* 71, 3659–3667. doi: 10.1128/AEM.71.7.3659-3667.2005

Degeest, B., and De Vuyst, L. (2000). Correlation of activities of the enzymes alpha-phosphoglucomutase, UDP-galactose 4-epimerase, and UDP-glucose pyrophosphorylase with exopolysaccharide biosynthesis by *Streptococcus thermophilus* LY03. *Appl. Environ. Microbiol.* 66, 3519–3527. doi: 10.1128/aem.66.8.3519-3527.2000

Delorme, C., Abraham, A. L., Renault, P., and Guédon, E. (2015). Genomics of *Streptococcus salivarius*, a major human commensal. *Infect. Genet. Evol.* 33, 381–392. doi: 10.1016/j.meegid.2014.10.001

Delorme, C., Bartholini, C., Bolotine, A., Ehrlich, S. D., and Renault, P. (2010). Emergence of a cell wall protease in the *Streptococcus thermophilus* population. *Appl. Environ. Microbiol.* 76, 451–460. doi: 10.1128/AEM.01018-09

Delorme, C., Bartholini, C., Luraschi, M., Pons, N., Loux, V., Almeida, M., et al. (2011). Complete genome sequence of the pigmented *Streptococcus thermophilus* strain JIM8232. *J. Bacteriol.* 193, 5581–5582. doi: 10.1128/JB.05404-11

Delorme, C., Legravet, N., Jamet, E., Hoarau, C., Alexandre, B., El-Sharoud, W. M., et al. (2017). Study of *Streptococcus thermophilus* population on a world-wide and historical collection by a new MLST scheme. *Int. J. Food Microbiol.* 242, 70–81. doi: 10.1016/j.ijfoodmicro.2016.11.016

Dupuis, M.-È., Villion, M., Magadán, A. H., and Moineau, S. (2013). CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. *Nat. Commun.* 4:2087. doi: 10.1038/ncomms3087

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Eisen, J. A., Heidelberg, J. F., White, O., and Salzberg, S. L. (2000). Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* 1:RESEARCH0011. doi: 10.1186/gb-2000-1-6-research 0011

Elbourne, L. D., Tetu, S. G., Hassan, K. A., and Paulsen, I. T. (2017). TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.* 45, D320–D324. doi: 10.1093/nar/gkw1068

Eng, C., Thibessard, A., Danielsen, M., Rasmussen, T. B., Mari, J. F., and Leblond, P. (2011). *In silico* prediction of horizontal gene transfer in *Streptococcus thermophilus*. *Arch. Microbiol.* 193, 287–297. doi: 10.1007/s00203-010-0671-8

Ercolini, D., Fusco, V., Blaiotta, G., and Coppola, S. (2005). Sequence heterogeneity in the lacSZ operon of *Streptococcus thermophilus* and its use in PCR systems for strain differentiation. *Res. Microbiol.* 156, 161–172. doi: 10.1016/j.resmic.2004.09.005

European Food Safety Authority [EFSA], (2007). Opinion of the scientific committee on a request from EFSA on the introduction of a Qualified Presumption of Safety (QPS) approach for assessment of selected microorganisms referred to EFSA. *EFSA J.* 587, 1–16. doi: 10.2903/j.efsa.2007.587

Evivie, S. E., Li, B., Ding, X., Meng, Y., Yu, S., Du, J., et al. (2017). Complete genome sequence of *Streptococcus thermophilus* KLDS 3.1003, a strain with high antimicrobial potential against foodborne and vaginal pathogens. *Front. Microbiol.* 8:1238. doi: 10.3389/fmicb.2017.01238

Fernandez-Gomez, B., Fernandez-Guerra, A., Casamayor, E. O., Gonzalez, J. M., Pedros-Alio, C., and Acinas, S. G. (2012). Patterns and architecture of genomic islands in marine bacteria. *BMC Genomics* 13:347. doi: 10.1186/1471-2164-13-347

Food and Drug Administration [FDA] (2007). *21 CFR Part 131: Microorganisms & Microbial-Derived Ingredients Used in Food (Partial List)*. Silver Spring, MA: FDA.

Galia, W., Jameh, N., Perrin, C., Genay, M., and Dary-Mourot, A. (2016). Acquisition of PrtS in *Streptococcus thermophilus* is not enough in certain strains to achieve rapid milk acidification. *Dairy Sci. Technol.* 96, 623–636. doi: 10.1007/s13594-016-0292-3

Galia, W., Perrin, C., Genay, M., and Dary, A. (2009). Variability and molecular typing of *Streptococcus thermophilus* strains displaying different proteolytic and acidifying properties. *Int. Dairy J.* 19, 89–95. doi: 10.1016/j.idairyj.2008.08.004

Garault, P., Letort, C., Juillard, V., and Monnet, V. (2000). Branched-chain amino acid biosynthesis is essential for optimal growth of *Streptococcus thermophilus* in milk. *Appl. Environ. Microbiol.* 66, 5128–5133. doi: 10.1128/aem.66.12.5128-5133.2000

Gardini, F., Rossi, F., Rizzotti, L., Torriani, S., Grazia, L., Chiavari, C., et al. (2012). Role of *Streptococcus thermophilus* PRI60 in histamine accumulation in cheese. *Int. Dairy J.* 27, 71–76. doi: 10.1016/j.idairyj.2012.07.005

Geertsma, E. R., Duurkens, R. H., and Poolman, B. (2005). The activity of the lactose transporter from *Streptococcus thermophilus* is increased by phosphorylated IIA and the action of β-galactosidase. *Biochemistry* 44, 15889–15897. doi: 10.1021/bi051638w

Giaretta, S., Treu, L., Vendramin, V., Da Silva Duarte, V., Tarrah, A., Campanaro, S., et al. (2018). Comparative transcriptomic analysis of *Streptococcus thermophilus* TH1436 and TH1477 showing different capability in the use of galactose. *Front. Microbiol.* 9:1765. doi: 10.3389/fmicb.2018.01765

Giraffa, G., Paris, A., Valcavi, L., Gatti, M., and Neviani, E. (2001). Genotypic and phenotypic heterogeneity of *Streptococcus thermophilus* strains isolated from dairy products. *J. Appl. Microbiol.* 91, 937–943. doi: 10.1046/j.1365-2672.2001.01464.x

Goh, Y. J., Goin, C., O'flaherty, S., Altermann, E., and Hutkins, R. (2011). Specialized adaptation of a lactic acid bacterium to the milk environment: the comparative genomics of *Streptococcus thermophilus* LMD-9. *Microb. Cell Fact.* 10(Suppl. 1):S22. doi: 10.1186/1475-2859-10-S1-S22

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360

Hafeez, Z., Cakir-Kiefer, C., Girardet, J.-M., Lecomte, X., Paris, C., Galia, W., et al. (2015). New insights into the proteolytic system of *Streptococcus thermophilus*: use of isracidin to characterize cell-associated extracellular peptidase activities. *J. Agric. Food Chem.* 63, 7522–7531. doi: 10.1021/acs.jafc.5b01647

Hafeez, Z., Cakir-Kiefer, C., Lecomte, X., Miclo, L., and Dary-Mourot, A. (2019). The X-prolyl dipeptidyl-peptidase PepX of *Streptococcus thermophilus* initially described as intracellular is also responsible for peptidase extracellular activity. *J. Dairy Sci.* 102, 113–123. doi: 10.3168/jds.2018-14823

Han, X., Yang, Z., Jing, X., Yu, P., Zhang, Y., Yi, H., et al. (2016). Improvement of the texture of yogurt by use of exopolysaccharide producing lactic acid bacteria. *Biomed. Res. Int.* 2016:7945675. doi: 10.1155/2016/7945675

Hao, M., Cui, Y., and Qu, X. (2018). Analysis of CRISPR-Cas system in *Streptococcus thermophilus* and its application. *Front. Microbiol.* 9:257. doi: 10.3389/fmicb.2018.00257

Hatmaker, E. A., Riley, L. A., O'dell, K. B., Papanek, B., Graveley, B. R., Garrett, S. C., et al. (2018). Complete genome sequence of industrial dairy strain *Streptococcus thermophilus* DGCC 7710. *Genome Announc.* 6:e01587-17. doi: 10.1128/genomeA.01587-17

Herve-Jimenez, L., Guillouard, I., Guedon, E., Boudebbouze, S., Hols, P., Monnet, V., et al. (2009). Postgenomic analysis of *Streptococcus thermophilus* cocultivated in milk with *Lactobacillus delbrueckii* subsp. *bulgaricus*: involvement of nitrogen, purine, and iron metabolism. *Appl. Environ. Microbiol.* 75, 2062–2073. doi: 10.1128/AEM.01984-08

Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., et al. (2005). New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. *FEMS Microbiol. Rev.* 29, 435–463. doi: 10.1016/j.fmrre.2005.04.008

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170. doi: 10.1126/science.1179555

Horvath, P., Romero, D. A., Coute-Monvoisin, A. C., Richards, M., Deveau, H., Moineau, S., et al. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1401–1412. doi: 10.1128/JB.01415-07

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, 680–682. doi: 10.1093/bioinformatics/btq003

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248

Iyer, R., Tomar, S. K., Uma Maheswari, T., and Singh, R. (2010). *Streptococcus thermophilus* strains: multifunctional lactic acid bacteria. *Int. Dairy J.* 20, 133–141. doi: 10.1016/j.idairyj.2009.10.005

Jans, C., Follador, R., Hochstrasser, M., Lacroix, C., Meile, L., and Stevens, M. J. A. (2013a). Comparative genome analysis of *Streptococcus infantarius* subsp. *infantarius* CJ18, an African fermented camel milk isolate with adaptations to dairy environment. *BMC Genomics* 14:200. doi: 10.1186/1471-2164-14-200

Jans, C., Kaindi, D. W. M., Böck, D., Njage, P. M. K., Kouamé-Sina, S. M., Bonfoh, B., et al. (2013b). Prevalence and comparison of *Streptococcus infantarius* subsp. *infantarius* and *Streptococcus gallolyticus* subsp. *macedonicus* in raw and fermented dairy products from East and West Africa. *Int. J. Food Microbiol.* 167, 186–195. doi: 10.1016/j.ijfoodmicro.2013.09.008

Juhas, M., Van Der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* 33, 376–393. doi: 10.1111/j.1574-6976.2008.00136.x

Junges, R., Maienschein-Cline, M., Morrison, D. A., and Petersen, F. C. (2019). Complete genome sequence of *Streptococcus pneumoniae* serotype 19F strain EF3030. *Microbiol. Resour. Announc.* 8:e00198-19. doi: 10.1128/MRA.00198-19

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016a). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kanehisa, M., Sato, Y., and Morishima, K. (2016b). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kang, X., Ling, N., Sun, G., Zhou, Q., Zhang, L., and Sheng, Q. (2012). Complete genome sequence of *Streptococcus thermophilus* strain MN-ZLW-002. *J. Bacteriol.* 194, 4428–4429. doi: 10.1128/JB.00740-12

Kelleher, P., Bottacini, F., Mahony, J., Kilcawley, K. N., and Van Sinderen, D. (2017). Comparative and functional genomics of the *Lactococcus lactis* Taxon; insights into evolution and niche adaptation. *BMC Genomics* 18:267. doi: 10.1186/s12864-017-3650-5

Kongo, J. M. (2013). "Lactic acid bacteria as starter-cultures for cheese processing: past, present and future developments," in *Lactic Acid Bacteria - R & D for Food, Health and Livestock Purposes*, ed. J. M. Kongo, (London: IntechOpen), doi: 10.5772/55937

Labrie, S. J., Tremblay, D. M., Plante, P. L., Wasserscheid, J., Dewar, K., Corbeil, J., et al. (2015). Complete genome sequence of *Streptococcus thermophilus* SMQ-301, a model strain for phage-host interactions. *Genome Announc.* 3:e00480-15. doi: 10.1128/genomeA.00480-15

Lefébure, T., and Stanhope, M. J. (2007). Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol.* 8:R71. doi: 10.1186/gb-2007-8-5-r71

Lerat, L., and Ochman, H. (2005). Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Res.* 33, 3125–3132. doi: 10.1093/nar/gki631

Letort, C., and Juillard, V. (2001). Development of a minimal chemically-defined medium for the exponential growth of *Streptococcus thermophilus*. *J. Appl. Microbiol.* 91, 1023–1029. doi: 10.1046/j.1365-2672.2001.01469.x

Li, B., Ding, X., Evivie, S. E., Jin, D., Meng, Y., Huo, G., et al. (2018). Short communication: genomic and phenotypic analyses of exopolysaccharides produced by *Streptococcus thermophilus* KLDS SM. *J. Dairy Sci.* 101, 106–112. doi: 10.3168/jds.2017-13534

Limauro, D., Falciatore, A., Basso, A. L., Forlani, G., and De Felice, M. (1996). Proline biosynthesis in *Streptococcus thermophilus*: characterization of the proBA operon and its products. *Microbiology* 142, 3275–3282. doi: 10.1099/13500872-142-11-3275

Linares, D. M., Arboleya, S., Ross, R. P., and Stanton, C. (2017). Complete genome sequence of the gamma-aminobutyric acid-producing strain *Streptococcus thermophilus* APC151. *Genome Announc.* 5:e00205-17. doi: 10.1128/genomeA.00205-17

Linares, D. M., O'callaghan, T. F., O'connor, P. M., Ross, R. P., and Stanton, C. (2016). *Streptococcus thermophilus* APC151 strain is suitable for the manufacture of naturally GABA-enriched bioactive yogurt. *Front. Microbiol.* 7:1876. doi: 10.3389/fmicb.2016.01876

Liu, M., Bayjanov, J. R., Renckens, B., Nauta, A., and Siezen, R. J. (2010). The proteolytic system of lactic acid bacteria revisited: a genomic comparison. *BMC Genomics* 11:36. doi: 10.1186/1471-2164-11-36

Liu, M., Siezen, R. J., and Nauta, A. (2009). *In silico* prediction of horizontal gene transfer events in *Lactobacillus bulgaricus* and *Streptococcus thermophilus* reveals protocooperation in yogurt manufacturing. *Appl. Environ. Microbiol.* 75, 4120–4129. doi: 10.1128/AEM.02898-08

Lluis-Arroyo, D., Flores-Nájera, A., Cruz-Guerrero, A., Gallardo-Escamilla, F., Lobato-Calleros, C., Jiménez-Guzmán, J., et al. (2014). Effect of an exopolysaccharide-producing strain of *Streptococcus thermophilus* on the yield and texture of Mexican Manchego-type cheese. *Int. J. Food Prop.* 17, 1680–1693. doi: 10.1080/10942912.2011.599091

Louis, E. P., Wei, Y., and Terns, M. P. (2017). Investigating the molecular mechanism of CRISPR-Cas adaptation of *Streptococcus thermophilus*. *J. Immunol.* 198(Suppl. 1):67.11.

Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., et al. (2006). Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15611–15616. doi: 10.1073/pnas.0607117103

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569

Manat, G., Roure, S., Auger, R., Bouhss, A., Barreteau, H., Mengin-Lecreulx, D., et al. (2014). Deciphering the metabolism of undecaprenyl-phosphate: the bacterial cell-wall unit carrier at the membrane frontier. *Microb. Drug Resist.* 20, 199–214. doi: 10.1089/mdr.2014.0035

Mayo, B., Van Sinderen, D., and Ventura, M. (2008). Genome analysis of food grade lactic Acid-producing bacteria: from basics to applications. *Curr. Genomics* 9, 169–183. doi: 10.2174/138920208784340731

Monnet, C., Mora, D., and Corrieu, G. (2005). Glutamine synthesis is essential for growth of *Streptococcus thermophilus* in milk and is linked to urea catabolism. *Appl. Environ. Microbiol.* 71, 3376–3378. doi: 10.1128/AEM.71.6.3376-3378.2005

Mora, D., Fortina, M. G., Parini, C., Ricci, G., Gatti, M., Giraffa, G., et al. (2002). Genetic diversity and technological properties of *Streptococcus thermophilus* strains isolated from dairy products. *J. Appl. Microbiol.* 93, 278–287. doi: 10.1046/j.1365-2672.2002.01696.x

Mora, D., Maguin, E., Masiero, M., Parini, C., Ricci, G., Manachini, P. L., et al. (2004). Characterization of urease genes cluster of *Streptococcus thermophilus*. *J. Appl. Microbiol.* 96, 209–219. doi: 10.1046/j.1365-2672.2003.02148.x

Mora, D., Monnet, C., Parini, C., Guglielmetti, S., Mariani, A., Pintus, P., et al. (2005). Urease biogenesis in *Streptococcus thermophilus*. *Res. Microbiol.* 156, 897–903. doi: 10.1016/j.resmic.2005.04.005

Moschetti, G., Blaiotta, G., Aponte, M., Catzeddu, P., Villani, F., Deiana, P., et al. (1998). Random amplified polymorphic DNA and amplified ribosomal DNA spacer polymorphism: powerful methods to differentiate *Streptococcus thermophilus* strains. *J. Appl. Microbiol.* 85, 25–36. doi: 10.1046/j.1365-2672.1998.00461.x

Ninova-Nikolova, N., and Urshev, Z. (2013). Fast acidifying urease-deficient *Streptococcus thermophilus* isolate shows spontaneous deletion of its complete urease operon. *Bulg. J. Agric. Sci.* 19, 112–116.

Nishimura, J., Kawai, Y., Aritomo, R., Ito, Y., Makino, S., Ikegami, S., et al. (2013). Effect of formic acid on exopolysaccharide production in skim milk fermentation by *Lactobacillus delbrueckii* subsp. *bulgaricus* OLL1073R-1. *Biosci. Microbiota Food Health* 32, 23–32. doi: 10.12938/bmfh.32.23

Padmanabhan, A., Tong, Y., Wu, Q., Zhang, J., and Shah, N. P. (2018). Transcriptomic insights into the growth phase- and sugar-associated changes in the exopolysaccharide production of a high EPS-producing *Streptococcus thermophilus* ASCC 1275. *Front. Microbiol.* 9:1919. doi: 10.3389/fmicb.2018.01919

Papadimitriou, K., Anastasiou, R., Maistrou, E., Plakas, T., Papandreou, N. C., Hamodrakas, S. J., et al. (2015a). Acquisition through horizontal gene transfer of plasmid pSMA198 by *Streptococcus macedonicus* ACA-DC 198 points towards the dairy origin of the species. *PLoS One* 10:e0116337. doi: 10.1371/journal.pone.0116337

Papadimitriou, K., Pot, B., and Tsakalidou, E. (2015b). How microbes adapt to a diversity of food niches. *Curr. Opin. Food Sci.* 2, 29–35. doi: 10.1016/j.cofs.2015.01.001

Papadimitriou, K., Anastasiou, R., Mavrogonatou, E., Blom, J., Papandreou, N. C., Hamodrakas, S. J., et al. (2014). Comparative genomics of the dairy isolate *Streptococcus macedonicus* ACA-DC 198 against related members of the *Streptococcus bovis*/*Streptococcus equinus* complex. *BMC Genomics* 15:272. doi: 10.1186/1471-2164-15-272

Pastink, M. I., Teusink, B., Hols, P., Visser, S., De Vos, W. M., and Hugenholtz, J. (2009). Genome-scale model of *Streptococcus thermophilus* LMG18311 for metabolic comparison of lactic acid bacteria. *Appl. Environ. Microbiol.* 75, 3627–3633. doi: 10.1128/AEM.00138-09

Pernoud, S., Fremaux, C., Sepulchre, A., Corrieu, G., and Monnet, C. (2004). Effect of the metabolism of urea on the acidifying activity of *Streptococcus thermophilus*. *J. Dairy Sci.* 87, 550–555. doi: 10.3168/jds.S0022-0302(04)73196-3

Proust, L., Loux, V., Martin, V., Magnabosco, C., Pedersen, M., Monnet, V., et al. (2018). Complete genome sequence of the industrial fast-acidifying strain

Streptococcus thermophilus N4L. *Microbiol. Resour. Announc.* 7:e01029-18. doi: 10.1128/MRA.01029-18

Purwandari, U., Shah, N. P., and Vasiljevic, T. (2007). Effects of exopolysaccharide-producing strains of *Streptococcus thermophilus* on technological and rheological properties of set-type yoghurt. *Int. Dairy J.* 17, 1344–1352. doi: 10.1016/j.idairyj.2007.01.018

Qiao, Y., Liu, G., Leng, C., Zhang, Y., Lv, X., Chen, H., et al. (2018). Metabolic profiles of cysteine, methionine, glutamate, glutamine, arginine, aspartate, asparagine, alanine and glutathione in *Streptococcus thermophilus* during pH-controlled batch fermentations. *Sci. Rep.* 8:12441. doi: 10.1038/s41598-018-30272-5

Rantsiou, K., Urso, R., Dolci, P., Comi, G., and Cocolin, L. (2008). Microflora of Feta cheese from four Greek manufacturers. *Int. J. Food Microbiol.* 126, 36–42. doi: 10.1016/j.ijfoodmicro.2008.04.031

Renye, J. A. Jr., Needleman, D. S., Somkuti, G. A., and Steinberg, D. H. (2017). Complete genome sequence of *Streptococcus thermophilus* strain B59671, which naturally produces the broad-spectrum bacteriocin thermophilin 110. *Genome Announc.* 5:e01213-17. doi: 10.1128/genomeA.01213-17

Renye, J. A. Jr., Needleman, D. S., and Steinberg, D. H. (2019). Complete genome sequences of bacteriocin-producing *Streptococcus thermophilus* strains ST106 and ST109. *Microbiol. Resour. Announc.* 8:e01336-18. doi: 10.1128/MRA.01336-18

Repar, J., and Warnecke, T. (2017). Non-random inversion landscapes in prokaryotic genomes are shaped by heterogeneous selection pressures. *Mol. Biol. Evol.* 34, 1902–1911. doi: 10.1093/molbev/msx127

Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2005). REBASE–restriction enzymes and DNA methyltransferases. *Nucleic Acids Res.* 33, D230–D232. doi: 10.1093/nar/gki029

Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 43, D298–D299. doi: 10.1093/nar/gku1046

Rodriguez-R, L. M., Gunturu, S., Harvey, W. T., Rosselló-Mora, R., Tiedje, J. M., Cole, J. R., et al. (2018). The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of *Archaea* and *Bacteria* at the whole genome level. *Nucleic Acids Res.* 46, W282–W288. doi: 10.1093/nar/gky467

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* 39, 9275–9282. doi: 10.1093/nar/gkr606

Savijoki, K., Ingmer, H., and Varmanen, P. (2006). Proteolytic systems of lactic acid bacteria. *Appl. Microbiol. Biotechnol.* 71, 394–406. doi: 10.1007/s00253-006-0427-1

Scott, E. J. II, Luke-Marshall, N. R., Campagnari, A. A., and Dyer, D. W. (2019). Draft genome sequence of pediatric otitis media isolate *Streptococcus pneumoniae* strain EF3030, which forms *in vitro* biofilms that closely mimic *in vivo* biofilms. *Microbiol. Resour. Announc.* 8:e01114-18. doi: 10.1128/MRA.01114-18

Selle, K., Klaenhammer, T. R., and Barrangou, R. (2015). CRISPR-based screening of genomic island excision events in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8076–8081. doi: 10.1073/pnas.1508525112

Settachaimongkon, S., Nout, M. J., Antunes Fernandes, E. C., Hettinga, K. A., Vervoort, J. M., Van Hooijdonk, T. C., et al. (2014). Influence of different proteolytic strains of *Streptococcus thermophilus* in co-culture with *Lactobacillus delbrueckii* subsp. *bulgaricus* on the metabolite profile of set-yoghurt. *Int. J. Food Microbiol.* 177, 29–36. doi: 10.1016/j.ijfoodmicro.2014.02.008

Shi, Y., Chen, Y., Li, Z., Yang, L., Chen, W., and Mu, Z. (2015). Complete genome sequence of *Streptococcus thermophilus* MN-BM-A02, a rare strain with a high acid-producing rate and low post-acidification ability. *Genome Announc.* 3:e00979-15. doi: 10.1128/genomeA.00979-15

Sieuwerts, S., Molenaar, D., Van Hijum, S. A., Beerthuyzen, M., Stevens, M. J., Janssen, P. W., et al. (2010). Mixed-culture transcriptome analysis reveals the molecular basis of mixed-culture growth in *Streptococcus thermophilus* and *Lactobacillus bulgaricus*. *Appl. Environ. Microbiol.* 76, 7775–7784. doi: 10.1128/AEM.01122-10

Solovyev, V., and Salamov, A. (2011). "Automatic annotation of microbial genomes and metagenomic sequences," in *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, ed. R. W. Li, (New York, NY: Nova Science Publishers), 61–78.

Song, X., Huang, H., Xiong, Z., Xia, Y., Wang, G., Yin, B., et al. (2018). Characterization of a cryptic plasmid isolated from *Lactobacillus casei* CP002616 and construction of shuttle vectors based on its Replicon. *J. Dairy Sci.* 101, 2875–2886. doi: 10.3168/jds.2017-13771

Sørensen, K. I., Curic-Bawden, M., Junge, M. P., Janzen, T., and Johansen, E. (2016). Enhancing the sweetness of yoghurt through metabolic remodeling of carbohydrate metabolism in *Streptococcus thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Appl. Environ. Microbiol.* 82, 3683–3692. doi: 10.1128/AEM.00462-16

Sujoy, B., and Aparna, A. (2013). Enzymology, immobilization and applications of urease enzyme. *Int. Res. J. Biol. Sci.* 2, 51–56.

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Sun, Z., Chen, X., Wang, J., Zhao, W., Shao, Y., Wu, L., et al. (2011). Complete genome sequence of *Streptococcus thermophilus* strain ND03. *J. Bacteriol.* 193, 793–794. doi: 10.1128/JB.01374-10

Tamulaitis, G., Kazlauskienë, M., Manakova, E., Venclovas, Č, Nwokeoji, A., Dickman, J., et al. (2014). Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. *Mol. Cell* 56, 506–517. doi: 10.1016/j.molcel.2014.09.027

Tian, H., Li, B., Evivie, S. E., Sarker, S. K., Chowdhury, S., Lu, J., et al. (2018). Technological and genomic analysis of roles of the cell-envelope protease PrtS in yoghurt starter development. *Int. J. Mol. Sci.* 19:E1068. doi: 10.3390/ijms19041068

Vaillancourt, K., Bedard, N., Bart, C., Tessier, M., Robitaille, G., Turgeon, N., et al. (2008). Role of galK and galM in galactose metabolism by *Streptococcus thermophilus*. *Appl. Environ. Microbiol.* 74, 1264–1267. doi: 10.1128/AEM.01585-07

Van Mastrigt, O., Di Stefano, E., Hartono, S., Abee, T., and Smid, E. J. (2018). Large plasmidome of dairy *Lactococcus lactis* subsp. *lactis* biovar diacetylactis FM03P encodes technological functions and appears highly unstable. *BMC Genomics* 19:620. doi: 10.1186/s12864-018-5005-2

Vaughan, E. E., Kleerebezem, M., and de Vos, W. M. (2003). "Genetics of the metabolism of lactose and other sugars," in *Genetics of Lactic Acid Bacteria*, eds B. J. B. Wood, and P. J. Warner, (New York, NY: Kluwer Academic), 95–119. doi: 10.1007/978-1-4615-7090-5_4

Vaughan, E. E., Van Den Bogaard, P. T., Catzeddu, P., Kuipers, O. P., and De Vos, W. M. (2001). Activation of silent *gal* genes in the *lac-gal* regulon of *Streptococcus thermophilus*. *J. Bacteriol.* 183, 1184–1194. doi: 10.1128/JB.183.4.1184-1194.2001

Wang, T., Xu, Z., Lu, S., Xin, M., and Kong, J. (2016). Effects of glutathione on acid stress resistance and symbiosis between *Streptococcus thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Int. Dairy J.* 61, 22–28. doi: 10.1016/j.idairyj.2016.03.012

Wassenaar, T. M., and Lukjancenko, O. (2014). "Comparative genomics of *Lactobacillus* and other LAB," in *Lactic Acid Bacteria: Biodiversity and Taxonomy*, eds W. H. Holzapfel, and B. J. B. Wood, (Hoboken, NJ: John Wiley & Sons), 55–69. doi: 10.1002/9781118655252.ch5

Wu, Q., and Shah, N. P. (2018). Comparative mRNA-Seq analysis reveals the improved EPS production machinery in *Streptococcus thermophilus* ASCC 1275 during optimized milk fermentation. *Front. Microbiol.* 9:445. doi: 10.3389/fmicb.2018.00445

Wu, Q., Tun, H. M., Leung, F. C., and Shah, N. P. (2014). Genomic insights into high exopolysaccharide-producing dairy starter bacterium *Streptococcus thermophilus* ASCC 1275. *Sci. Rep.* 4:4974. doi: 10.1038/srep04974

Xiong, Z.-Q., Kong, L.-H., Lai, P. F. H., Xia, Y.-J., Liu, J.-C., Li, Q.-Y., et al. (2019a). Genomic and phenotypic analyses of exopolysaccharide biosynthesis in *Streptococcus thermophilus* S-3. *J. Dairy Sci.* 102, 4925–4934. doi: 10.3168/jds.2018-15572

Xiong, Z.-Q., Kong, L.-H., Meng, H.-L., Cui, J.-M., Xia, Y.-J., Wang, S.-J., et al. (2019b). Comparison of gal-lac operons in wild-type galactose-positive and -negative *Streptococcus thermophilus* by genomics and transcription analysis. *J. Ind. Microbiol. Biotechnol.* 46, 751–758. doi: 10.1007/s10295-019-02145-x

Yamauchi, R., Maguin, E., Horiuchi, H., Hosokawa, M., and Sasaki, Y. (2019). The critical role of urease in yogurt fermentation with various combinations of *Streptococcus thermophilus* and *Lactobacillus delbrueckii* ssp. *bulgaricus*. *J. Dairy Sci.* 102, 1033–1043. doi: 10.3168/jds.2018-15192

Yu, J., Sun, Z., Liu, W., Xi, X., Song, Y., Xu, H., et al. (2015). Multilocus sequence typing of *Streptococcus thermophilus* from naturally fermented dairy foods in China and Mongolia. *BMC Microbiol.* 15:236. doi: 10.1186/s12866-015-0551-0

Zotta, T., Ricciardi, A., Ciocia, F., Rossano, R., and Parente, E. (2008). Diversity of stress responses in dairy thermophilic streptococci. *Int. J. Food Microbiol.* 124, 34–42. doi: 10.1016/j.ijfoodmicro.2008.02.024

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4

The complete genome sequence of the yogurt isolate *Streptococcus thermophilus* ACA-DC 2

Alexandraki et al. Standards in Genomic Sciences (2017) 12:18
DOI 10.1186/s40793-017-0227-5

Standards in Genomic Sciences

EXTENDED GENOME REPORT Open Access

The complete genome sequence of the yogurt isolate *Streptococcus thermophilus* ACA-DC 2

Voula Alexandraki[1], Maria Kazou[1], Jochen Blom[2], Bruno Pot[3], Effie Tsakalidou[1] and Konstantinos Papadimitriou[1*]

Abstract

Streptococcus thermophilus ACA-DC 2 is a newly sequenced strain isolated from traditional Greek yogurt. Among the 14 fully sequenced strains of *S. thermophilus* currently deposited in the NCBI database, the ACA-DC 2 strain has the smallest chromosome, containing 1,731,838 bp. The annotation of its genome revealed the presence of 1,850 genes, including 1,556 protein-coding genes, 70 RNA genes and 224 potential pseudogenes. A large number of pseudogenes were identified. This was also accompanied by the absence of pathogenic features suggesting evolution of strain ACA-DC 2 through genome decay processes, most probably due to adaptation to the milk ecosystem. Analysis revealed the existence of one complete lactose-galactose operon, several proteolytic enzymes, one exopolysaccharide cluster, stress response genes and four putative antimicrobial peptides. Interestingly, one CRISPR-cas system and one orphan CRISPR, both carrying only one spacer, were predicted indicating low activity or inactivation of the cas proteins. Nevertheless, four putative restriction-modification systems were determined that may compensate any deficiencies of the CRISPR-cas system. Furthermore, whole genome phylogeny indicated three distinct clades within *S. thermophilus*. Comparative analysis among selected strains representative for each clade, including strain ACA-DC 2, revealed a high degree of conservation at the genomic scale, but also strain specific regions. Unique genes and genomic islands of strain ACA-DC 2 contained a number of genes potentially acquired through horizontal gene transfer events, that could be related to important technological properties for dairy starters. Our study suggests genomic traits in strain ACA-DC 2 compatible to the production of dairy fermented foods.

Keywords: Extended genome report, *Streptococcus thermophilus*, Yogurt, Horizontal gene transfer, CRISPR, Stress genes

Introduction

The use of microorganisms in food fermentations is the means for converting perishable and frequently inedible raw materials into safe, shelf-stable and nutritionally upgraded foods [1]. The economic importance of starter cultures for the food industry has led to the continuous search for the discovery of new microorganisms with important technological characteristics. In many cases it has been proven that traditionally fermented foods represent a natural reservoir of undiscovered microbial strains for possible diverse food applications [2, 3].

Streptococcus thermophilus is among the species commonly used in the dairy industry, mainly in the fermentation of yogurt and several cheese varieties, contributing to the desirable organoleptic characteristics of the final product [4, 5]. It is the sole species considered GRAS within the *Streptococcus* genus, which includes mostly pathogens and opportunistic pathogens [6]. Due to the industrial significance of the species, a plethora of studies has been conducted for a number of strains, revealing information about their diverse technological features [7, 8]. Furthermore, during the last 15 years, the advance of high-throughput sequencing techniques along with the development of novel

* Correspondence: kpapadimitriou@aua.gr
[1]Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Iera Odos 75, 118 55 Athens, Greece
Full list of author information is available at the end of the article

bioinformatics tools facilitated the analysis of complete genome sequences, providing information for the overall genetic content of *S. thermophilus* [9–12]. These studies have demonstrated that *S. thermophilus* strains have been adapted to the milk environment through extensive reductive evolution as indicated by the large number of pseudogenes found in all strains. Adaptation to the milk environment is also supported by the loss of genes related to carbohydrate metabolism and virulence.

In this study, we present the analysis of the complete genome sequence of *S. thermophilus* ACA-DC 2. The genomic insights acquired could be proven useful for the exploitation of the specific strain in the production of fermented dairy products.

Organism information

Classification and features

Streptococcus thermophilus ACA-DC 2 is classified within the order *Lactobacillales* of the class *Bacilli*. It is a non-sporulating, Gram-positive bacterium with coccus-shaped cells (Fig. 1). The strain was isolated from traditional Greek yogurt manufactured through backslopping [13, 14]. Its optimum growth takes place in M17 medium at 42 °C under microaerophilic conditions within 24 h. Information about the classification and the features of *S. thermophilus* ACA-DC 2 is summarized in Table 1. The phylogenetic analysis was based on 16S rRNA gene sequences and places *S. thermophilus* ACA-DC 2 in the distinct cluster formed by the *S. thermophilus* strains and within the salivarius group, as shown in Fig. 2.

Genome sequencing information

Genome project history

S. thermophilus ACA-DC 2 is deposited in the ACA-DC culture collection of the Laboratory of Dairy Research, Agricultural University of Athens, Athens, Greece. The strain was selected for sequencing in order to obtain information about its technological and probiotic potential, having as basic aim its application as a starter culture in the production of dairy fermented foods. The project was carried out in 2015 and the genome was sequenced, fully assembled and annotated. The genome sequencing project was registered in the European Nucleotide Archive database under accession number LT604076. The summary of the project information is shown in Table 2.

Growth conditions and genomic DNA preparation

S. thermophilus ACA-DC 2 was grown in M17 medium (Biokar Diagnostics, Beauvais, France). For the isolation of the genomic DNA, 2 ml from an overnight culture incubated at 42 °C were used and the extraction procedure was performed according to the protocol of Pitcher et al. [15]. The purity and the concentration of the extracted DNA were measured with a UV-Vis spectrophotometer (Q5000, Quawell, San Jose, USA) while its integrity was evaluated electrophoretically in a 0.8% agarose gel.

Genome sequencing and assembly

Whole-genome sequencing was performed using the Illumina HiSeq2500 and the PacBio RSII platforms at BaseClear service laboratory for DNA-research (Leiden, The Netherlands). Paired-end sequence reads were generated using the Illumina HiSeq2500 system. FASTQ sequence files were obtained using the Illumina Casava pipeline v1.8.3. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing adapters and/or PhiX control signal were removed using an in-house filtering protocol. The second quality assessment was based on the remaining reads using the FASTQC quality control tool v0.10.0 resulting in 4,403,680 reads.

The data collected from the PacBio RSII instrument were processed and filtered using the SMRT Analysis software suite. The Continuous Long Read data were

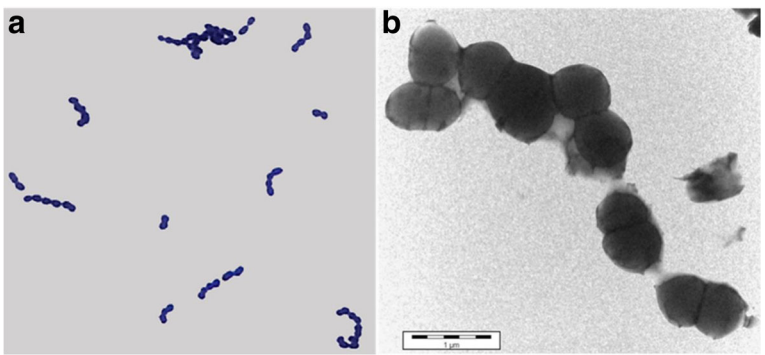


Fig. 1 Photomicrographs of *S. thermophilus* ACA-DC 2. The images were obtained with (**a**) optical microscopy at magnification 1000x for Gram stained cells and (**b**) transmission electron microscopy of stained cells with 10% (w/v) PTA. Scale bar in (**b**) corresponds to 1 μm

Table 1 Classification and general features of *S. thermophilus* strain ACA-DC 2 according to the MIGS recommendations [39]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS [40]
		Phylum *Firmicutes*	TAS [41, 42]
		Class *Bacilli*	TAS [43, 44]
		Order *Lactobacillales*	TAS [44, 45]
		Family *Streptococcaceae*	TAS [46–48]
		Genus *Streptococcus*	TAS [47, 49, 50]
		Species *Streptococcus thermophilus*	TAS [47, 51, 52]
		Strain: ACA-DC 2	TAS (this study)
	Gram stain	Positive	IDA
	Cell shape	Coccus	IDA
	Motility	Non-motile	IDA
	Sporulation	Non-sporulating	NAS
	Temperature range	30–50 °C	TAS [51]
	Optimum temperature	42 °C	TAS [53]
	pH range; Optimum	5–7; 6.5	TAS [53]
	Carbon source	lactose; saccharose; d-glucose; galactose	IDA
MIGS-6	Habitat	Yogurt	TAS [13, 14]
MIGS-6.3	Salinity	2% NaCl (w/v)	TAS [51]
MIGS-22	Oxygen requirement	Microaerophilic	TAS [51]
MIGS-15	Biotic relationship	Free-living	NAS
MIGS-14	Pathogenicity	Non-pathogen	NAS
MIGS-4	Geographic location	Greece	TAS [13, 14]
MIGS-5	Sample collection	1988	NAS
MIGS-4.1	Latitude	Unknown	
MIGS-4.2	Longitude	Unknown	
MIGS-4.4	Altitude	Unknown	

[a]Evidence codes - *IDA* inferred from direct assay, *TAS* traceable author statement (i.e., a direct report exists in the literature), *NAS* non-traceable author statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [54]

filtered by Read-length (>50), Subread-length (>50) and Read quality (>0.75) resulting in 117,020 reads.

The quality of the Illumina FASTQ sequences was enhanced by trimming off low-quality bases using the program bbduk, which is part of the BBMap suite v34.46. The quality-filtered sequence reads were puzzled into a number of contig sequences. The analysis was performed using ABySS v1.5.1. The contigs were linked and placed into super-scaffolds based on the alignment of the PacBio CLR reads with BLASR [16]. The alignment was further used to estimate the orientation, order and distance between the contigs by the SSPACE-LongRead scaffolder v1.0 [17]. The gapped regions within the super-scaffolds were closed in an automated manner using GapFiller v1.10 [18]. The method takes advantage of the insert size between the Illumina paired-end reads. The assembly resulted in one circular chromosome of 1,731,838 bp.

Genome annotation

Prediction of genes was carried out with the online programs Prodigal [19], MetaGeneAnnotator [20] and FGENESB [21], for comparison and verification of the obtained results. Genome annotation was performed using RAST v2.0 [22]. Annotation anomalies, including pseudogenes, were identified using GenePRIMP [23]. All data acquired were combined and subjected to manual curation. The WebMGA server [24] and the EggNog v4.5 [25] were used for COG annotation, the Phobius web server was used for the identification of genes with transmembrane helices and genes with signal peptides [26] and the Pfam database was used for the identification of genes with Pfam domains [27]. Potential pathogenic features were identified using the MP3 tool [28]. The CRISPRs, the restriction-modification systems and the putative antimicrobial peptides were predicted using the CRISPRFinder web tool [29], the REBASE database [30] and BAGEL3 [31], respectively. The KODON software (Applied Maths NV, Sint-Martens-Latem, Belgium) was utilized for the visualization of synteny among the CRISPR regions of ACA-DC 2 and LMD-9 strains. The EDGAR server [32] was used for whole genome phylogeny and Venn diagram analysis. Circoletto [33] was employed for whole genome alignment among *S. thermophilus* strains. Finally, the genomic islands were identified through the IslandViewer 3 web-based resource [34].

Genome properties

The complete genome of *S. thermophilus* ACA-DC 2 consists of one circular chromosome containing 1,731,838 bp. The average GC content of the chromosome is 39.2%. A total of 1,850 genes were predicted after manual curation, including 1,556 protein-coding genes, 70 RNAs (56 tRNAs and 14 rRNAs) and 224 potential pseudogenes (Table 3). A circular map of the genome was generated using the CGView comparison tool [35] as shown in Fig. 3. Function was assigned to 1,182 genes (63.89%), while 1,318 genes (71.24%) had one or more conserved Pfam domains. The distribution of protein-coding genes into COG

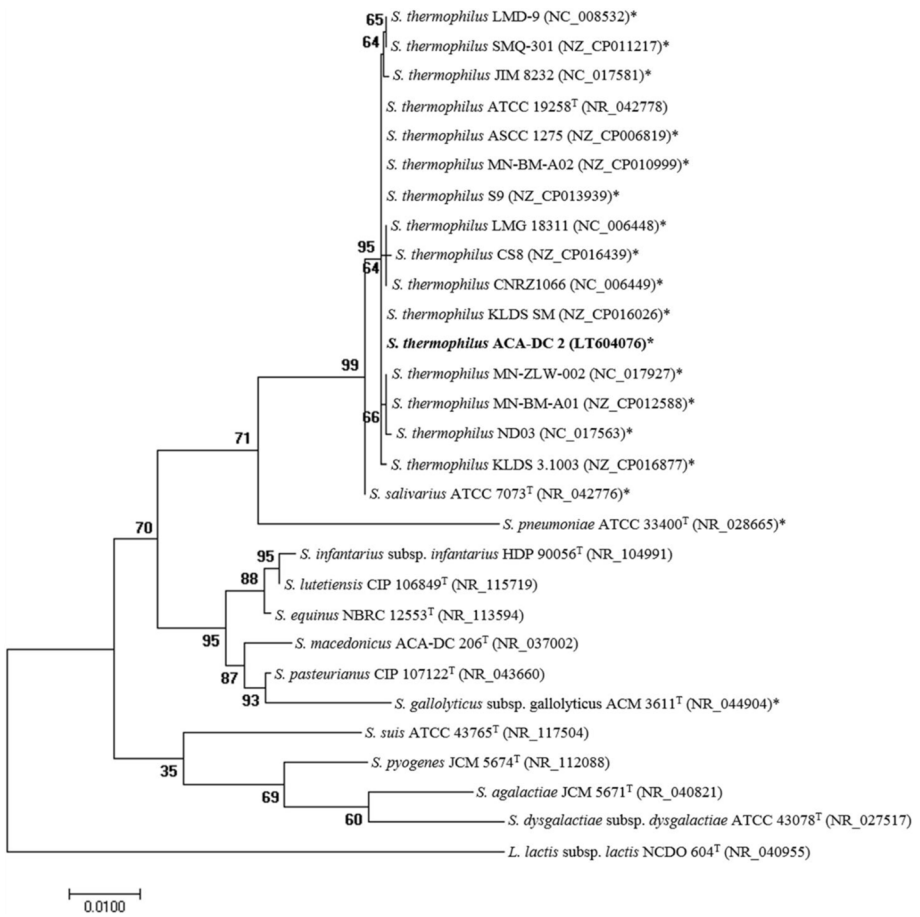


Fig. 2 Phylogenetic tree highlighting the position of *S. thermophilus* ACA-DC 2 relative to other *Streptococcus* species. The tree was constructed based on 16S rRNA gene sequences. GenBank accession numbers are presented in parentheses and type strains are indicated with a superscript T (type strains = ^{T}). Strains with complete genome sequence are marked with an asterisk. 16S rRNA gene sequences were aligned using MUSCLE [55]. The phylogenetic tree was built by the Maximum Likelihood method within MEGA7 software [56] using the Tamura-Nei substitution model [57]. *Lactococcus lactis* subsp. *lactis* NCDO 604^{T} served as the outgroup. Bootstrap values derived after 1,000 replicates. The scale bar indicates an estimated 0.01 nucleotide change per nucleotide position

functional category is shown in Table 4. The analysis revealed that approximately 28.5% of the protein-coding genes do not have any described function.

Insights from the genome sequence

Main genome sequence characteristics

The genome of *S. thermophilus* ACA-DC 2 is the smallest one described to date among the fully sequenced strains of the species deposited in NCBI and it is approximately 200 kbp smaller than the larger described genome. The majority of potential pseudogenes encode hypothetical proteins, transposases and proteins involved in carbohydrate transport and metabolism. Analysis of the genome for virulence factors with the MP3 tool revealed a number of hits (data not shown). Detailed inspection of these hits with EDGAR demonstrated that several such genes are conserved among *S. thermophilus* strains indicating that it is rather unlikely to be related to pathogenicity, given the safe history of the species. The high percentage of pseudogenes along with the absence of typical virulence factors for streptococci suggest that the ACA-DC 2 strain evolved through genome decay during the adaptation to the rich in nutrients dairy environment [9, 11].

S. thermophilus ACA-DC 2 carries a complete lactose-galactose operon containing the *galR*, *galK*, *galT*, *galE*, *galM*, *lacS* and *lacZ* genes (STACADC2_1195-1189) and it is able to ferment lactose and galactose, the latter in a fairly slow rate (data not shown). It has been reported that fermentation of galactose is limited among the strains of *S. thermophilus* [11]. As mentioned above,

Table 2 Project information

MIGS ID	Property	Term
MIGS 31	Finishing quality	Finished
MIGS-28	Libraries used	Illumina genomic Nextera XT library; PacBio 10 kb genomic library
MIGS 29	Sequencing platforms	Illumina HiSeq2500; PacBio RSII
MIGS 31.2	Fold coverage	636x
MIGS 30	Assemblers	ABySS v1.5.1; BLASR; SSPACE v1.0; GapFiller v1.10
MIGS 32	Gene calling method	Prodigal; MeteGeneAnnotator; FGENESB
	Locus Tag	STACADC2
	Genbank ID	LT604076
	GenBank Date of Release	29-Jul-2016
	GOLD ID	NA
	BIOPROJECT	PRJEB14916
MIGS 13	Source Material Identifier	ACA-DC 2
	Project relevance	Dairy isolate

several genes responsible for the transport and degradation of sugars, such as fructose, maltose and trehalose, have been identified as pseudogenes in the genome of ACA-DC 2, further supporting the specialization of the bacterium to catabolize lactose.

The proteolytic system of *S. thermophilus* ACA-DC 2 consists of several genes encoding aminopeptidases, such as *pepA* (STACADC2_1626), *pepC* (STACADC2_0202), *pepF* (STACADC2_0406), *pepM* (STACADC2_1333), *pepN* (STACADC2_0892), *pepO* (STACADC2_1656), *pepP* (STACADC2_1520), *pepQ* (STACADC2_0572), *pepS* (STACADC2_0058), *pepT* (STACADC2_0971), *pepV* (STACADC2_0960), and *pepX* (STACADC2_1446), one oligopeptide *opp* ABC transporter (STACADC2_1229-1233), four polar amino acid ABC transporters (STACADC2_0780-0782, STACADC2_0992-0995, STACADC2_1355-1358, STACADC2_1431-1433), two symporters for branched-chain amino acids (STACADC2_0872, STACADC2_1160), and two glutamine ABC transporters (STACADC2_0547-0548, STACADC2_1281-1282). Strain ACA-DC 2 lacks a cell wall-associated proteinase (PrtS). Although this gene may be important for optimal growth of *S. thermophilus* in milk, its absence is of minor significance when co-cultured with a proteolytic *Lactobacillus delbrueckii* subsp. *bulgaricus* strain, since the release of peptides by the latter is beneficial for the growth of *S. thermophilus* [10, 11].

Similar to other dairy bacteria, *S. thermophilus* ACA-DC 2 is able to synthesize exopolysaccharides (EPS) that may confer improved viscosity and texture to yogurt [4]. The EPS cluster is flanked by a *deoD* gene encoding a purine nucleoside phosphorylase (STACADC2_0949) and a pseudogene originally encoding a beta-glucosidase. Four of these genes, namely *epsA* (STACADC2_0948), *epsB* (STACADC2_0947), *epsC* (STACADC2_0946) and *epsD* (STACADC2_0945) are implicated in the regulation, polymerization, chain length and export of the EPS and are conserved among several *S. thermophilus* strains [36].

The genome analysis of strain ACA-DC 2 revealed a number of genes known to be responsive to unfavorable conditions prevailing during industrial applications. Among them we identified conserved heat shock genes like *grp*E, *dna*K, *dna*J (STACADC2_0105-0107) and *gro*ES, *gro*EL (STACADC2_0179-0180), genes encoding Clp proteases (STACADC2_0071, STACADC2_0315, STACADC2_0526, STACADC2_0544, STACADC2_1391), a proton translocating F_0F_1-ATPase system (STACADC2_0430-0437) and a P-type calcium pump ATPase (STACADC2_0983). The strain also harbors genes related to oxidative stress, namely a Mn-superoxide dismutase (STACADC2_0657), a glutathione reductase (STACADC2_0362), two thioredoxins (STACADC2_1043, STACADC2_1624), two thioredoxin reductases (STACADC2_1208, STACADC2_1429), a NADH oxidase (STACADC2_1113) and two sulfoxide reductases (STACADC2_1408, STACADC2_1159). Furthermore, the genome carries four putative antimicrobial peptides that need further investigation (STACADC2_0091, STACADC2_1453, STACADC2_1458 and STACADC2_1709).

Table 3 Genome statistics

Attribute	Value	% of Total
Genome size (bp)	1,731,838	100.00
DNA coding (bp)	1,356,670	78.34
DNA G + C (bp)	679,104	39.21
DNA scaffolds	1	100.00
Total genes	1,850	100.00
Protein coding genes	1,556	84.11
RNA genes	70	3.78
Pseudo genes	224	12.11
Genes in internal clusters	NA	NA
Genes with function prediction	1,182	63.89
Genes assigned to COGs	1,327	71.73
Genes with Pfam domains	1,318	71.24
Genes with signal peptides	127	6.86
Genes with transmembrane helices	339	18.32
CRISPR repeats	2	

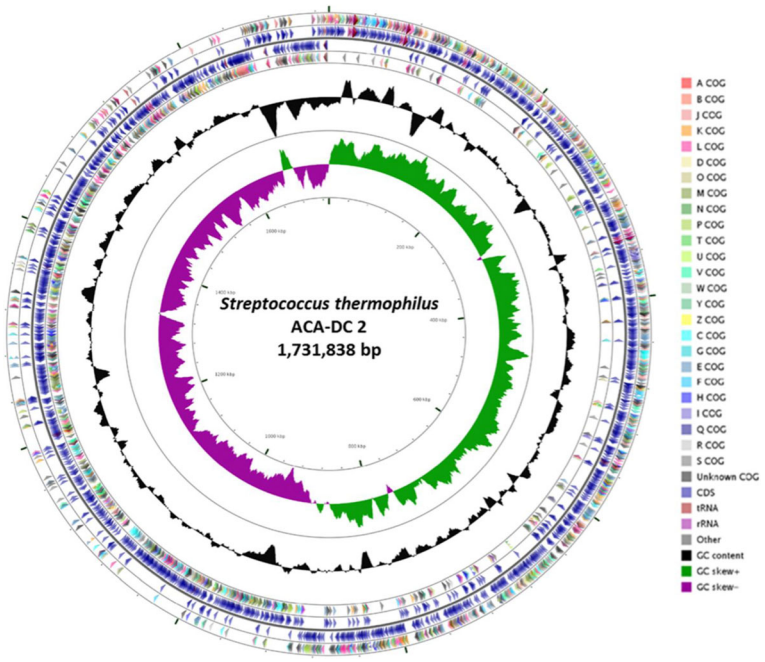


Fig. 3 Circular map of *S. thermophilus* ACA-DC 2 genome features generated with the CGview tool. From periphery to center: Protein coding genes on forward strand colored by COG category assignment; Genes on forward strand; Protein coding genes on reverse strand colored by COG category assignment; Genes on reverse strand; GC content; GC skew; Genome region in kbp

Two candidate CRISPRs were detected in the chromosome of strain ACA-DC 2. Intriguingly, both CRISPRs contained only one spacer. One CRISPR was found surrounded by cas proteins (STACADC2_0849-0856) while the other was orphan. The CRISPR-cas system of strain ACA-DC 2 exhibited the same organization and high degree of identity to that described previously for strain LMD-9 (Fig. 4) [37]. The two CRISPR-cas systems differed mainly in the *csm6* gene, which in the case of strain ACA-DC 2 is a potential pseudogene as well as in *csm2* gene that seems to be distinct in the two strains. *S. thermophilus* LMD-9 carries three CRISPR-cas systems and the system that is similar to that of ACA-DC 2 carries the lowest number (three) of spacers. Combined these findings could indicate low activity or even inactivation of the entire CRISPR-cas system in strain ACA-DC 2. Another possibility that cannot be excluded concerns low exposure of strain ACA-DC 2 to foreign DNA. Of course, any deficiency in the activity of the CRISPR-cas system may be compensated by restriction-modification (RM) systems. Strain ACA-DC 2 carries four putative RM systems according to the REBASE database (data not shown) belonging to RM types I (STACADC2_0642, STACADC2_0645, STACADC2_0648), II (STACADC2_0597-0598), III (STACADC2_0788-0789) and IV (STACADC2_0626).

Comparative genomic analysis, strain specific genomic features and genomic islands

Resolution of phylogenetic trees based on 16S rRNA gene sequences is limited due to high sequence identity, especially for strains of the same species. For this reason, we also performed whole genome phylogeny as implemented in EDGAR, using all available complete genomes of *S. thermophilus*. The phylogenetic tree produced revealed that *S. thermophilus* strains could be clustered in two distinct branches, the second of which could be also split in two sub-branches (Fig. 5). Strain ACA-DC 2 formed one of the branches along with strains CNRZ1066, LMG 18311, S9 and CS8. We chose strains ACA-DC 2, JIM 8232 and KLDS 3.1003 as representatives of each branch for comparative genomic analysis (Fig. 5). Whole genome alignments revealed extensive regions of high identity (>98%) among the genomes. However, regions of lower identity (between 80 and 98%) as well as strain specific regions were also identified. Using Venn diagram analysis as implemented in EDGAR, we determined a core genome of 1,303 genes among the three genomes as well as 137, 185 and 236 unique genes for strains ACA-DC 2, KLDS 3.1003 and JIM 8232, respectively.

The 137 unique genes of strain ACA-DC 2 were found to be involved in diverse functions (Fig. 6). At least some of those genes may be the result of horizontal gene

Table 4 Number of genes associated with general COG functional categories

Code	Value	%age	Description
J	146	9.38	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	89	5.72	Transcription
L	136	8.74	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	16	1.03	Cell cycle control, Cell division, chromosome partitioning
V	39	2.51	Defense mechanisms
T	43	2.76	Signal transduction mechanisms
M	80	5.14	Cell wall/membrane biogenesis
N	3	0.19	Cell motility
U	20	1.29	Intracellular trafficking and secretion
O	55	3.53	Posttranslational modification, protein turnover, chaperones
C	40	2.57	Energy production and conversion
G	66	4.24	Carbohydrate transport and metabolism
E	160	10.28	Amino acid transport and metabolism
F	67	4.31	Nucleotide transport and metabolism
H	49	3.15	Coenzyme transport and metabolism
I	33	2.12	Lipid transport and metabolism
P	67	4.31	Inorganic ion transport and metabolism
Q	13	0.84	Secondary metabolites biosynthesis, transport and catabolism
R	63	4.05	General function prediction only
S	215	13.82	Function unknown
-	229	14.72	Not in COGs

The total is based on the total number of protein coding genes in the genome

transfer (HGT). HGT acquired genes may play a role in the technological properties of *S. thermophilus* strains [11]. Another analysis that may also reveal regions of HGT in the bacterial chromosome is the identification of GIs [38]. Twelve integrated GIs were predicted in the genome of *S. thermophilus* ACA-DC 2 (Fig. 6), containing a total of 213 genes also involved in diverse functions (Fig. 6). Detailed analysis of genes either unique or in the GIs could relate some of them to important technological traits. For example, we determined genes coding for cold shock proteins CspA and CspG (STACADC2_0749-0750), acid resistance locus arl7 (STACADC2_0743), putative bacteriocin peptides (STACADC2_1453 and STACADC2_1458) and a type I RM system (STACADC2_0642, STACADC2_0645, STACADC2_0648). A putative agmatinase gene (STACADC2_0818) that may play a role to protocooperation of *S. thermophilus* and *L. bulgaricus* during polyamine metabolism, was also detected in ACA-DC 2 strain [10]. Furthermore, genes implicated in zinc and heavy metals transport (STACADC2_0165-0166, STACADC2_0752), in DNA repair and metabolism (STACADC2_1696, STACADC2_1716, STACADC2_1719, STACADC2_1754) as well as several ribosome binding proteins, were also identified (STACADC2_0137, STACADC2_1568-1569, STACADC2_1667, STACADC2_1669-1671, STACADC2_1675-1695, STACADC2_1717, STACADC2_1732-1733, STACADC2_1752, STACADC2_1755).

Conclusions

The genome of *S. thermophilus* ACA-DC 2 presents characteristics in accordance with its adaptation to the milk environment including a high percentage of pseudogenes and absence of pathogenic features. Our analysis revealed that the strain carries genes involved in lactose and galactose catabolism and protein degradation that may accommodate its growth during milk fermentation. Stress response related genes that may contribute to survival under technological hurdles were also detected. Whole genome phylogeny suggested that *S. thermophilus* strains may diversify in three phylogenetic clades. Comparative analysis of genomes representative of each clade, including strain ACA-DC 2, revealed a number of unique genes for the latter. Furthermore, certain unique genes or genes belonging to GIs could be related to technological

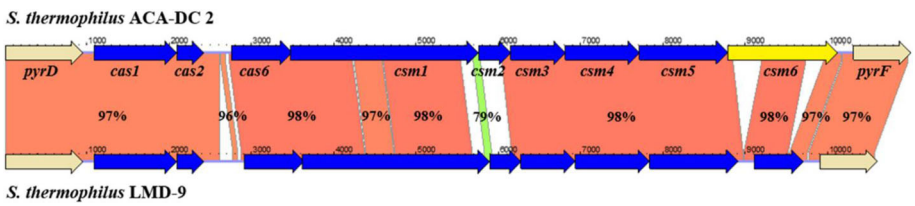


Fig. 4 Synteny plot of the CRISPR loci between *S. thermophilus* strains ACA-DC 2 and LMD-9. The synteny of the two regions was calculated by the KODON software. In both strains the *cas* genes are denoted in blue. Gene *csm6* in strain ACA-DC 2 is a potential pseudogene and it is denoted in yellow. The *pyrD* and *pyrF* genes colored in beige define the upstream and downstream limits of the CRISPR loci. Percentages displayed in the ribbon areas correspond to the % identity among the nucleotide sequences

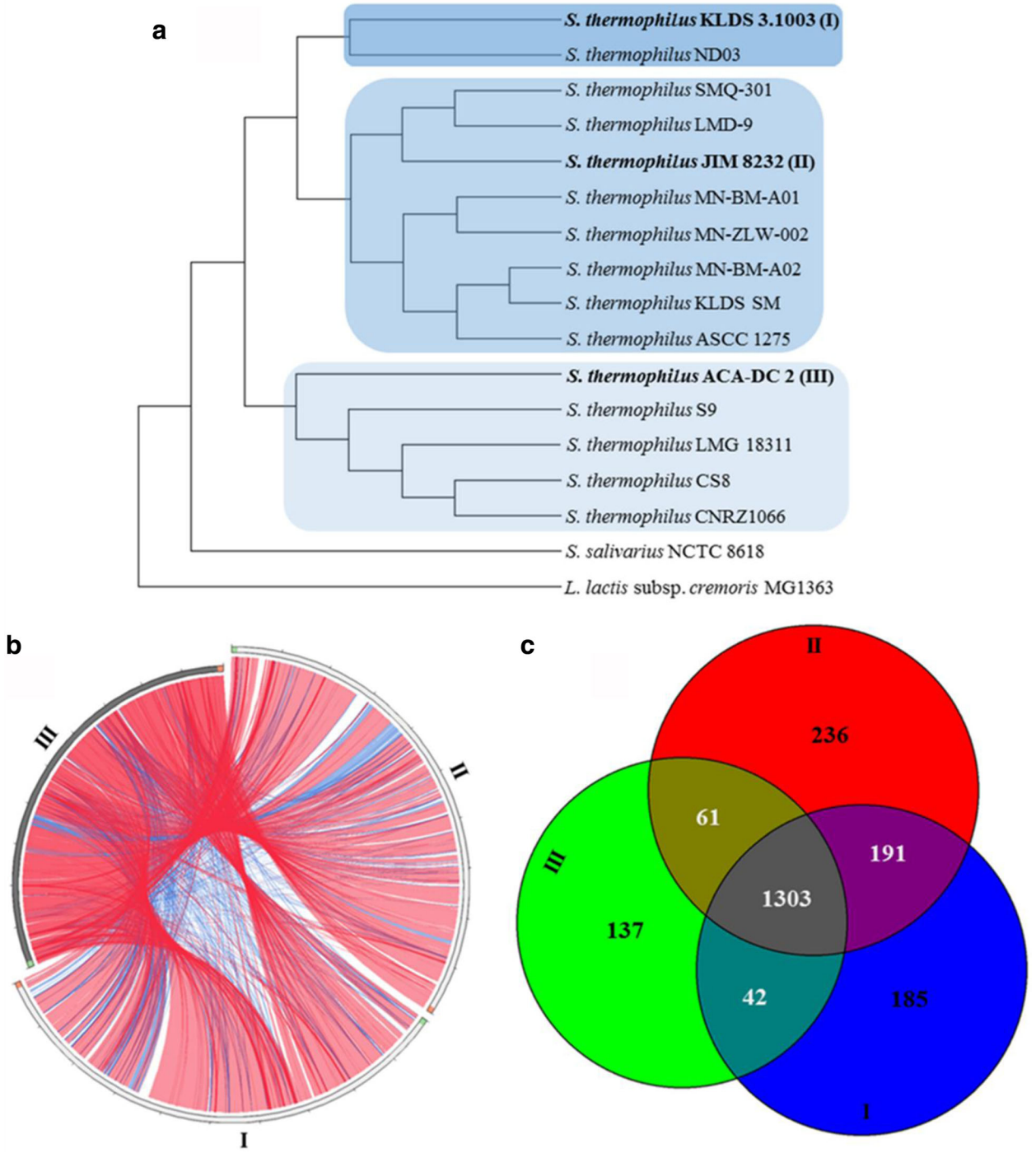


Fig. 5 Comparative genomics of *S. thermophilus* strains. **a** Whole genome phylogeny of *S. thermophilus* strains with complete available genome sequences. The phylogenetic tree was calculated in EDGAR and it is presented as a cladogram ignoring branch length. The strains *S. salivarius* NCTC 8616 and *Lactococcus lactis* subsp. *cremoris* MG1363 were used as outgroups. Colored boxes indicate the three distinct *S. thermophilus* branches identified. Strains in each colored branch designated in bold were used for further comparative analysis. **b** Whole genome alignments of *S. thermophilus* strains KLDS 3.1003, JIM 8232 and ACA-DC 2 using Circoletto. Red and blue ribbons correspond to regions of >98% and 80–98% identity, respectively. White regions correspond to strain specific loci. **c** Venn diagram analysis of *S. thermophilus* strains KLDS 3.1003, JIM 8232 and ACA-DC 2, as implemented in EDGAR

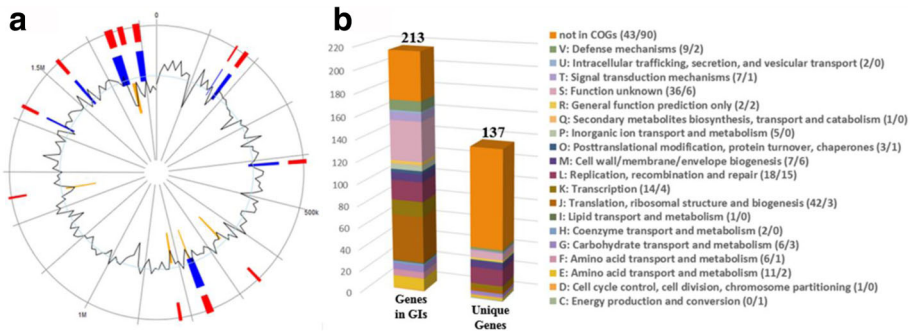


Fig. 6 Additional genomic features of *S. thermophilus* ACA-DC 2. **a** Circular map of the *S. thermophilus* ACA-DC 2 genome as generated by IslandViewer 3. Highlighted regions correspond to GIs. GIs are colored within the circular map according to the prediction method used: five GIs in orange and eight GIs in blue were predicted with SIGI-HMM and IslandPath-DIMOB, respectively. Twelve integrated GIs are presented on the periphery in red. The *black* line plot represents the GC content (%) of the genomic sequence. **b** Distribution of genes in GIs and unique genes of *S. thermophilus* ACA-DC 2 into COG categories

properties important for starter cultures. Theoretically, such genes could have been acquired through HGT. These findings render *S. thermophilus* ACA-DC 2 an appropriate candidate for use in the production of fermented dairy products.

Abbreviations

ACA-DC: Agricultural College of Athens - Dairy Collection; CLR: Continuous long read; COG: Clusters of Orthologous Groups; CRISPR: Clustered regularly interspaced short palindromic repeats; ENA: European nucleotide archive; EPS: Exopolysaccharide; GenePRIMP: Gene prediction improvement pipeline; GI: Genomic Island; GRAS: Generally regarded as safe; HGT: Horizontal gene transfer; PTA: Phosphotungstic acid; RAST: Rapid annotation using subsystem technology; RM: Restriction-modification; SMRT: Single molecule real time

Acknowledgements

We would like to thank Dr. Nikos Kyrpides at the Joint Genome Institute (United States Department of Energy) for the analysis of the ACA-DC 2 genome with the GenePrimp server. Furthermore we would like to thank Prof. Konstantinos Fasseas at the Faculty of Crop Science (Agricultural University of Athens) for the transmission electron microscopy experiment.

Funding

The present work was co-financed by the European Social Fund and the National resources EPEAEK and YPEPTH through the Thales project.

Authors' contributions

VA performed genome analysis and participated in the writing of the manuscript. MK performed genome analysis and participated in the writing of the manuscript. JB performed genome analysis. BP performed genome analysis. ET characterized strain ACA-DC 2, conceived the project and participated in the writing of the manuscript. KP conceived the project, performed genome analysis and participated in the writing of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Iera Odos 75, 118 55 Athens, Greece. [2]Bioinformatics & Systems Biology, Justus-Liebig-University Giessen, 35392 Giessen, Hesse, Germany. [3]Research Group of Industrial Microbiology and Food Biotechnology (IMDO), Vrije Universiteit Brussel, Brussels, Belgium.

Received: 9 November 2016 Accepted: 4 January 2017
Published online: 31 January 2017

References

1. Papadimitriou K, Pot B, Tsakalidou E. How microbes adapt to a diversity of food niches. Curr Opin Food Sci. 2015;2:29–35.
2. Poornachandra Rao K, Chennappa G, Suraj U, Nagaraja H, Charith Raj AP, Sreenivasa MY. Probiotic potential of *Lactobacillus* strains isolated from sorghum-based traditional fermented food. Probiotics Antimicrob Proteins. 2015;7:146–56.
3. Topisirovic L, Kojic M, Fira D, Golic N, Strahinic I, Lozo J. Potential of lactic acid bacteria isolated from specific natural niches in food production and preservation. Int J Food Microbiol. 2006;112:230–5.
4. Broadbent JR, McMahon DJ, Welker DL, Oberg CJ, Moineau S. Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. J Dairy Sci. 2003;86:407–23.
5. Leroy F, De Vuyst L. Lactic acid bacteria as functional starter cultures for the food fermentation industry. Trends Food Sci Tech. 2004;15:67–78.
6. Krzysciak W, Pluskwa KK, Jurczak A, Koscielniak D. The pathogenicity of the *Streptococcus* genus. Eur J Clin Microbiol Infect Dis. 2013;32:1361–76.
7. Iyer R, Tomar SK, Uma Maheswari T, Singh R. *Streptococcus thermophilus* strains: multifunctional lactic acid bacteria. Int Dairy J. 2010;20:133–41.
8. Rossi F, Marzotto M, Cremonese S, Rizzotti L, Torriani S. Diversity of *Streptococcus thermophilus* in bacteriocin production; inhibitory spectrum and occurrence of thermophilin genes. Food Microbiol. 2013;35:27–33.
9. Bolotin A, Quinquis B, Renault P, Sorokin A, Ehrlich SD, Kulakauskas S, et al. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat Biotechnol. 2004;22:1554–8.
10. Goh YJ, Goin C, O'Flaherty S, Altermann E, Hutkins R. Specialized adaptation of a lactic acid bacterium to the milk environment: the comparative genomics of *Streptococcus thermophilus* LMD-9. Microb Cell Fact. 2011;10(Suppl 1):S22.
11. Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, Leblond-Bourget N, et al. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. FEMS Microbiol Rev. 2005;29:435–63.
12. Prajapati JB, Nathani NM, Patel AK, Senan S, Joshi CG. Genomic analysis of dairy starter culture *Streptococcus thermophilus* MTCC 5461. J Microbiol Biotechnol. 2013;23:459–66.
13. Tsakalidou E, Manolopoulou E, Kabaraki E, Zoidou E, Pot B, Kersters K, et al. The combined use of whole cell protein extracts for the identification (SDS-PAGE) and enzyme activity screening of lactic acid bacteria isolated from traditional Greek dairy products. System Appl Microbiol. 1994;17:444–58.
14. Tsakalidou E, Zoidou E, Kalantzopoulos G. SDS-Polyacrylamide gel electrophoresis of cell proteins from *Lactobacillus delbreuckii* subsp. *bulgaricus* and *Streptococcus salivarius* subsp. *thermophilus* strains isolated from yoghurt and cheese. Milchwissenschaft. 1992;47:296–8.
15. Pitcher DG, Saunders NA, Owen RJ. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett Appl Microbiol. 1989;8:151–6.
16. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics. 2012;13:238.

17. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics. 2014;15:211.
18. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol. 2012;13:R56.
19. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
20. Noguchi H, Taniguchi T, Itoh T. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res. 2008;15:387–96.
21. Solovyev V, Salamov A. Automatic annotation of microbial genomes and metagenomic sequences. In: Li RW, editor. Metagenomics and its applications in agriculture, biomedicine and environmental studies. New York: Nova Science Publishers; 2011. p. 61–78.
22. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
23. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, et al. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods. 2010;7:455–7.
24. Wu S, Zhu Z, Fu L, Niu B, Li W. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 2011;12:444.
25. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 2016;44:D286–93.
26. Kall L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. Nucleic Acids Res. 2007;35:W429–32.
27. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279–85.
28. Gupta A, Kapil R, Dhakan DB, Sharma VK. MP3: a software tool for the prediction of pathogenic proteins in genomic and metagenomic data. PLoS One. 2014;9:e93907.
29. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35:W52–7.
30. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. 2010;38:D234–6.
31. van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, Kuipers OP. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. Nucleic Acids Res. 2013; 41:W448–53.
32. Blom J, Kreis J, Spanig S, Juhre T, Bertelli C, Ernst C, et al. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nucleic Acids Res. 2016;44:W22–8.
33. Darzentas N. Circoletto: visualizing sequence similarity with Circos. Bioinformatics. 2010;26:2620–1.
34. Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, et al. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Res. 2015;43:W104–8.
35. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 2008;36:W181–4.
36. Sun Z, Chen X, Wang J, Zhao W, Shao Y, Wu L, et al. Complete genome sequence of *Streptococcus thermophilus* strain ND03. J Bacteriol. 2011;193:793–4.
37. Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, et al. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. J Bacteriol. 2008;190:1401–12.
38. Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev. 2009;33:376–93.
39. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. Nat Biotechnol. 2008;26:541–7.
40. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576–9.
41. Gibbons NE, Murray RGE. Proposals concerning the higher taxa of Bacteria. Int J Syst Bacteriol. 1978;28:1–6.
42. Garrity G, Holt J. The road map to the manual. In: Boone D, Castenholz R, Garrity G, editors. Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer; 2001. p. 119–66.
43. Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, et al., editors. Bergey's manual of systematic bacteriology. vol. 3. 2nd ed. New York: Springer; 2009; 19–20.
44. Euzéby J. List of new names and new combinations previously effectively, but not validly, published. List no. 132. Int J Syst Evol Microbiol. 2010;60: 469–72.
45. Ludwig W, Schleifer KH, Whitman WB. Order II. *Lactobacillales* ord. nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, et al., editors. Bergey's manual of systematic bacteriology. vol. 3. 2nd ed. New York: Springer; 2009;464.
46. Goodfellow M, Kämpfer P, De Vos P, Rainey F, Schleifer KH, Whitman WB. Family VI. *Streptococcaceae* fam. nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, et al., editors. Bergey's manual of systematic bacteriology. vol. 3. 2nd ed. New York: Springer; 2009;655.
47. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol. 1980;30:225–420.
48. Deibel RH, Seeley HW. Family II. *Streptococcaceae*. In: Buchanan RE, Gibbons NE, editors. Bergey's manual of determinative bacteriology. 8th ed. Baltimore: The Williams and Wilkins Co; 1974. p. 490–515.
49. Whiley RA, Hardie JM. Genus I. *Streptococcus* gen. nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, et al., editors. Bergey's manual of systematic bacteriology. vol. 3. 2nd ed. New York: Springer; 2009. p. 655–710.
50. Rosenbach FJ. In: Bergmann JF, editor. Microorganismen bei den Wund-Infections-Krankheiten des Menschen. Wiesbaden: JF Bergmann; 1884;1–122.
51. Schleifer KH, Ehrmann M, Krusch U, Neve H. Revival of the species *Streptococcus thermophilus* (ex Orla-Jensen, 1919) nom. rev. System App Microbiol. 1991;14:386–8.
52. Orla-Jensen S. The lactic acid bacteria. 1919. p. 1–196.
53. Beal C, Louvet P, Corrieu G. Influence of controlled pH and temperature on the growth and acidification of pure cultures of *Streptococcus thermophilus* 404 and *Lactobacillus bulgaricus* 398. Appl Microbiol Biotechnol. 1989;32:148–54.
54. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25:25–9.
55. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
56. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4.
57. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993;10:512–26.

5

Complete genome sequence of the yogurt isolate *Lactobacillus delbrueckii subsp. bulgaricus* ACA-DC 87

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Complete Genome Sequence of the Yogurt Isolate *Lactobacillus delbrueckii* subsp. *bulgaricus* ACA-DC 87

Voula Alexandraki,[a] Maria Kazou,[a] Bruno Pot,[b] Effie Tsakalidou,[a] Konstantinos Papadimitriou[a]

Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Athens, Greece[a]; Research Group of Industrial Microbiology and Food Biotechnology (IMDO), Vrije Universiteit Brussel, Brussels, Belgium[b]

ABSTRACT *Lactobacillus delbrueckii* subsp. *bulgaricus* is widely used in the production of yogurt and cheese. In this study, we present the complete genome sequence of *L. delbrueckii* subsp. *bulgaricus* ACA-DC 87 isolated from traditional Greek yogurt. Whole-genome analysis may reveal desirable technological traits of the strain for dairy fermentations.

Received 11 July 2017 **Accepted** 12 July 2017 **Published** 24 August 2017

Citation Alexandraki V, Kazou M, Pot B, Tsakalidou E, Papadimitriou K. 2017. Complete genome sequence of the yogurt isolate *Lactobacillus delbrueckii* subsp. *bulgaricus* ACA-DC 87. Genome Announc 5:e00868-17. https://doi.org/10.1128/genomeA.00868-17.

Address correspondence to Konstantinos Papadimitriou, kpapadimitriou@aua.gr.

Lactobacillus delbrueckii subsp. *bulgaricus* is among the most important starters employed in dairy fermentations and is used mainly in the production of yogurt in protocooperation with *Streptococcus thermophilus* (1). *L. delbrueckii* subsp. *bulgaricus* belongs to the acidophilus complex, a group of lactobacilli which further includes *Lactobacillus acidophilus*, *Lactobacillus helveticus*, *Lactobacillus johnsonii*, and *Lactobacillus gasseri* (2). Here, we report the genome sequencing and the genomic features of *L. delbrueckii* subsp. *bulgaricus* ACA-DC 87, which was isolated from traditional Greek yogurt (3, 4).

The genome was sequenced at the Beijing Genomics Institute (BGI Co., Ltd., Hong Kong). A total of 17,362,540 paired-end reads (500-bp, 2,000-bp, and 6,000-bp libraries) were generated using the Illumina HiSeq 2000 platform with >100-fold sequence coverage. After filtering, the reads were assembled with SOAPdenovo version 1.05, resulting in one circular chromosome (5). The accuracy of the assembly was evaluated through whole-genome alignment of the ACA-DC 87 genome against a reference genome, namely, that of *L. delbrueckii* subsp. *bulgaricus* ATCC 11842 (6), using ProgressiveMauve (7). The ACA-DC 87 genome sequence was analyzed using 3 gene prediction programs, i.e., Prodigal (8), MetaGeneAnnotator (9), and FGENESB (10). Additionally, RAST version 2.0 was used both in genome annotation and prediction of rRNA and tRNA genes (11). The GenePRIMP pipeline was also employed for the detection of annotation anomalies, including putative pseudogenes (12). All the results obtained were manually curated. Further bioinformatics analysis focused on the identification of genomic islands (GIs) with IslandViewer 4 (13), clustered regularly interspaced short palindromic repeats (CRISPRs) with CRISPRFinder (14), and Pfam domain-containing proteins with Pfam database (15), while the WebMGA server was used for clusters of orthologous groups (COG) functional annotation (16).

The chromosome of ACA-DC 87 comprises 1,856,003 bp with a G+C content of 49.8%. A total of 1,993 genes were identified in the genome sequence, including 1,644 protein-coding genes and 229 potential pseudogenes. The genome also contains 93 tRNA genes and 9 complete rRNA operons. Twelve genomic islands (GIs) with a total of 196 genes were predicted in the genome of ACA-DC 87. These genes may have been acquired through horizontal gene transfer and could be related to technological features. Several of these genes encode CRISPR-associated proteins, subunits of

restriction-modification systems, and proteins implicated in exopolysaccharide biosynthesis. The genome also carries one confirmed CRISPR array with a size of 761 bp, containing 11 spacers. Furthermore, 1,322 proteins with Pfam domains were identified. The COG annotation results revealed that approximately 87% of the protein-coding genes (1,284 proteins) were assigned to at least one functional category. The higher number of proteins were allocated to the category of translation, ribosomal structure and biogenesis (J: 8.5%), followed by the categories of amino acid transport and metabolism (E: 7.8%) and replication, recombination, and repair (L: 7.4%). Further analysis of the ACA-DC 87 genome sequence may decipher the technological potential of the strain toward its application in the production of fermented dairy products.

Accession number(s). The genome sequence of *L. delbrueckii* subsp. *bulgaricus* ACA-DC 87 was deposited at the European Nucleotide Archive under the accession number LT899687.

ACKNOWLEDGMENTS

We thank Nikos Kyrpides at the Joint Genome Institute (United States Department of Energy) for analysis of the ACA-DC 87 genome with the GenePRIMP pipeline.

The present work was cofinanced by the European Social Fund and the national resources EPEAEK and YPEPTH through the Thales project.

REFERENCES

1. Angelov M, Kostov G, Simova E, Beshkova D, Koprinkova-Hristova P. 2009. Proto-cooperation factors in yogurt starter cultures. Rev Genie Indust 3:4–12.
2. Zhang ZG, Ye ZQ, Yu L, Shi P. 2011. Phylogenomic reconstruction of lactic acid bacteria: an update. BMC Evol Biol 11:1. https://doi.org/10.1186/1471-2148-11-1.
3. Tsakalidou E, Manolopoulou E, Kabaraki E, Zoidou E, Pot B, Kersters K, Kalantzopoulos G. 1994. The combined use of whole-cell protein extracts for the identification (SDS-PAGE) and enzyme activity screening of lactic acid bacteria isolated from traditional Greek dairy products. Syst Appl Microbiol 17:444–458. https://doi.org/10.1016/S0723-2020(11)80062-7.
4. Tsakalidou E, Zoidou E, Kalantzopoulos G. 1992. SDS-polyacrylamide gel electrophoresis of cell proteins from *Lactobacillus delbreuckii* subsp. *bulgaricus* and *Streptococcus salivarius* subsp. *thermophilus* strains isolated from yoghurt and cheese. Milchwissenschaft 47:296–298.
5. Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligonucleotide alignment program. Bioinformatics 24:713–714. https://doi.org/10.1093/bioinformatics/btn025.
6. van de Guchte M, Penaud S, Grimaldi C, Barbe V, Bryson K, Nicolas P, Robert C, Oztas S, Mangenot S, Couloux A, Loux V, Dervyn R, Bossy R, Bolotin A, Batto JM, Walunas T, Gibrat JF, Bessières P, Weissenbach J, Ehrlich SD, Maguin E. 2006. The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. Proc Natl Acad Sci U S A 103:9274–9279. https://doi.org/10.1073/pnas.0603024103.
7. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147. https://doi.org/10.1371/journal.pone.0011147.
8. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
9. Noguchi H, Taniguchi T, Itoh T. 2008. MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res 15:387–396. https://doi.org/10.1093/dnares/dsn027.
10. Solovyev V, Salamov A. 2011. Automatic annotation of microbial genomes and metagenomic sequences, p 61–78. *In* Li RW (ed), Metagenomics and its applications in agriculture, biomedicine and environmental studies. Nova Science Publishers, New York, NY.
11. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
12. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods 7:455–457. https://doi.org/10.1038/nmeth.1457.
13. Bertelli C, Laird MR, Williams KP; Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL, Brinkman FSL. 2 May 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale data sets. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx343.
14. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35:W52–W57. https://doi.org/10.1093/nar/gkm360.
15. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344.
16. Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: a customizable Web server for fast metagenomic sequence analysis. BMC Genomics 12:444. https://doi.org/10.1186/1471-2164-12-444.

Supporting information

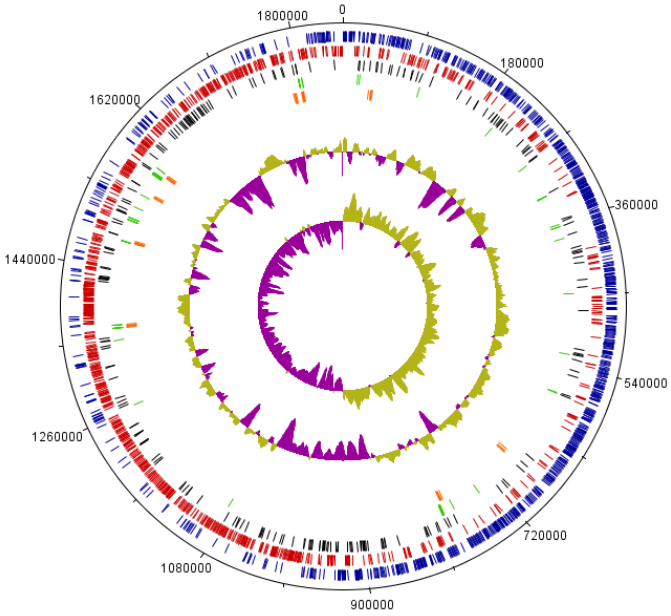


Figure 5.1 Circular map of the *Lb. bulgaricus* ACA-DC 87 genome. Each ring represents specific genomic features appearing from the periphery to the center of the map: Forward CDSs (blue); Reverse CDSs (red); Pseudogenes (black); tRNA (green); rRNA (orange); %GC plot; GC skew.

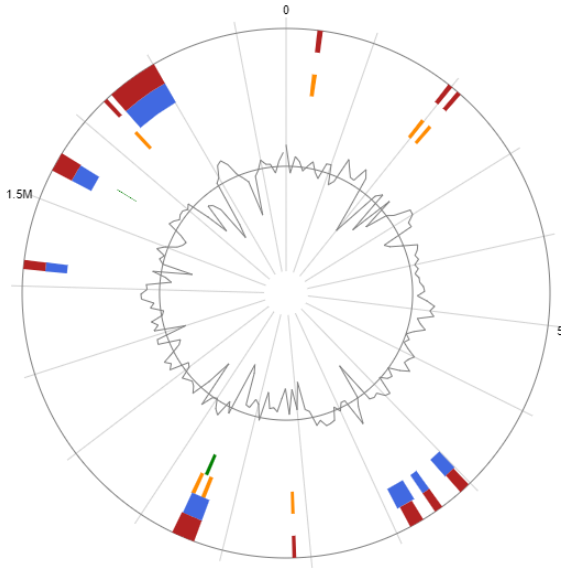


Figure 5.2 Genomic islands (GIs) of *Lb. bulgaricus* ACA-DC 87 as predicted by the IslandViewer 4. GIs are colored within the circular map based on the tools used, i.e. orange for SIGI-HMM, blue for IslandPath-DIMOB, green for IslandPick and red for an integration of the three tools employed. The prediction of the GIs is based on a hidden Markov model (SIGI-HMM), on features associated with genomic islands (IslandPath-DIMOB) and on comparative genomics (IslandPick). The black line represents the G+C content (%) of *Lb. bulgaricus* ACA-DC 87.

Table 5.1 Genes in the integrated genomic islands identified in the genome of *Lb. bulgaricus* ACA-DC 87

Genomic island	Locus_tag	Product
GI_1	LBUL87_0035	hypothetical protein
	LBUL87_0036	hypothetical protein
	LBUL87_0037	
GI_2	LBUL87_0196	
	LBUL87_0197	Dihydroneopterin aldolase
	LBUL87_0198	2-amino-4-hydroxy-6-hydroxymethyldihydropteridinepyrophosphokinase / GTP cyclohydrolase I type 1
	LBUL87_0199	Dihydrofolate synthase @ Folylpolyglutamate synthase
	LBUL87_0200	Dihydropteroate synthase
	LBUL87_0201	Adenosylhomocysteinase
	LBUL87_0202	
	LBUL87_0203	hypothetical protein
GI_3	LBUL87_0213	Catalyzes the cleavage of p-aminobenzoyl-glutamate to p-aminobenzoate and glutamate, subunit A
	LBUL87_0214	
	LBUL87_0215	
	LBUL87_0216	contains ATPase involved in DNA replication initiation domain
GI_4	LBUL87_0719	2',3'-cyclic-nucleotide 2'-phosphodiesterase
	LBUL87_0720	DNA helicase, phage-associated
	LBUL87_0721	FIG00746507: hypothetical protein
	LBUL87_0722	membrane protein
	LBUL87_0723	Hypothetical protein ywlG
	LBUL87_0724	
	LBUL87_0725	Pyrroline-5-carboxylate reductase
	LBUL87_0726	Gamma-glutamyl phosphate reductase
	LBUL87_0727	
	LBUL87_0728	hypothetical protein
	LBUL87_0729	DNA polymerase III alpha subunit
GI_5	LBUL87_0760	hypothetical protein
	LBUL87_0761	Protein gp28 [Bacteriophage A118]
	LBUL87_0762	Fumarate reductase flavoprotein subunit
	LBUL87_0763	GTP pyrophosphokinase
	LBUL87_0764	Regulation of D-alanyl-lipoteichoic acid biosynthesis, DltR
	LBUL87_0765	Regulation of D-alanyl-lipoteichoic acid biosynthesis, sensor histidine kinase
	LBUL87_0766	CRISPR-associated protein, Csn1 family
	LBUL87_0767	CRISPR-associated protein Cas1
	LBUL87_0768	CRISPR-associated protein
GI_6	LBUL87_0782	hypothetical protein
	LBUL87_0783	hypothetical protein
	LBUL87_0784	FIG00750780: hypothetical protein
	LBUL87_0785	3-oxoacyl-[acyl-carrier-protein] synthase, KASIII
	LBUL87_0786	Acyl carrier protein
	LBUL87_0787	Malonyl CoA-acyl carrier protein transacylase
	LBUL87_0788	3-oxoacyl-[acyl-carrier protein] reductase
	LBUL87_0789	3-oxoacyl-[acyl-carrier-protein] synthase, KASII

	LBUL87_0790	Biotin carboxyl carrier protein of acetyl-CoA carboxylase
	LBUL87_0791	3-hydroxyacyl-[acyl-carrier-protein] dehydratase, FabZ form
	LBUL87_0792	Biotin carboxylase of acetyl-CoA carboxylase
	LBUL87_0793	Acetyl-coenzyme A carboxyl transferase beta chain
	LBUL87_0794	Acetyl-coenzyme A carboxyl transferase alpha chain
	LBUL87_0795	Enoyl-[acyl-carrier-protein] reductase [NADH]
	LBUL87_0796	Biotin-protein ligase
	LBUL87_0797	FIG00743625: hypothetical protein
	LBUL87_0798	FIG00751483: hypothetical protein
	LBUL87_0799	FIG00747771: hypothetical protein
	LBUL87_0800	Mobile element protein
	LBUL87_0801	hypothetical protein
GI_7	LBUL87_0927	
	LBUL87_0928	
	LBUL87_0929	
	LBUL87_0930	hypothetical protein
	LBUL87_0931	hypothetical protein
	LBUL87_0932	hypothetical protein
GI_8	LBUL87_1050	FIG00742116: hypothetical protein
	LBUL87_1051	Tripeptide aminopeptidase
	LBUL87_1052	FIG137478: Hypothetical protein YbgI
	LBUL87_1053	Putative tRNA-m1A22 methylase
	LBUL87_1054	Mobile element protein
	LBUL87_1055	putative mutator protein
	LBUL87_1056	putative enterolysin A
	LBUL87_1057	Methylated-DNA--protein-cysteine methyltransferase
	LBUL87_1058	YeeC-like protein
	LBUL87_1059	McrBC 5-methylcytosine restriction system component
	LBUL87_1060	ATPas
	LBUL87_1061	hypothetical protein
	LBUL87_1062	FIG00748645: hypothetical protein
	LBUL87_1063	FIG00754632: hypothetical protein
	LBUL87_1064	Type I restriction-modification system, restriction subunit R
	LBUL87_1065	
	LBUL87_1066	Type I restriction-modification system, specificity subunit S
	LBUL87_1067	Type I restriction-modification system, specificity subunit S
	LBUL87_1068	Type I restriction-modification system, specificity subunit S
	LBUL87_1069	Type I restriction-modification system, DNA-methyltransferase subunit M
	LBUL87_1070	Type I restriction-modification system, DNA-methyltransferase subunit M
	LBUL87_1071	
	LBUL87_1072	Lead, cadmium, zinc and mercury transporting ATPase; Copper-translocating P-type ATPase
	LBUL87_1073	hypothetical protein
GI_9	LBUL87_1425	FIG00745302: hypothetical protein
	LBUL87_1426	Putative transposase (fragment)
	LBUL87_1427	FIG00752523: hypothetical protein
	LBUL87_1428	hypothetical protein

	LBUL87_1429	FmtA protein involved in methicillin resistance / affects cell wall cross-linking and amidation
	LBUL87_1430	
	LBUL87_1431	FmtA protein involved in methicillin resistance / affects cell wall cross-linking and amidation
	LBUL87_1432	DNA-binding protein HBsu
	LBUL87_1433	
	LBUL87_1434	
	LBUL87_1435	
	LBUL87_1436	Thiamin biosynthesis lipoprotein ApbE
	LBUL87_1437	
	LBUL87_1438	
	LBUL87_1439	
	LBUL87_1440	LSU ribosomal protein L7/L12 (P1/P2)
GI_10	LBUL87_1544	Putative regulator of the mannose operon, ManO
	LBUL87_1545	PTS system, mannose-specific IID component
	LBUL87_1546	PTS system, mannose-specific IIC component
	LBUL87_1547	PTS system, mannose-specific IIB component / PTS system, mannose-specific IIA component
	LBUL87_1548	FIG00749320: hypothetical protein
	LBUL87_1549	Glycosyltransferase
	LBUL87_1550	
	LBUL87_1551	Mobile element protein
	LBUL87_1552	Na+/H+ antiporter
	LBUL87_1553	hypothetical protein
	LBUL87_1554	
	LBUL87_1555	
	LBUL87_1556	Iron-sulfur cluster assembly protein SufB
	LBUL87_1557	Putative iron-sulfur cluster assembly scaffold protein for SUF system, SufE2
	LBUL87_1558	Cysteine desulfurase
	LBUL87_1559	Iron-sulfur cluster assembly protein SufD
	LBUL87_1560	Iron-sulfur cluster assembly ATPase protein SufC
GI_11	LBUL87_1639	
	LBUL87_1640	capsular polysaccharide biosynthesis protein
	LBUL87_1641	Glycosyltransferase
	LBUL87_1642	capsular polysaccharide biosynthesis protein
	LBUL87_1643	Glycosyltransferase
	LBUL87_1644	
GI_12	LBUL87_1650	FIG00751362: hypothetical protein
	LBUL87_1651	
	LBUL87_1652	
	LBUL87_1653	Transcriptional regulator, XRE family
	LBUL87_1654	Glycerol-3-phosphate dehydrogenase [NAD(P)+]
	LBUL87_1655	FIG00745125: hypothetical protein
	LBUL87_1656	
	LBUL87_1657	FIG007491: hypothetical protein YeeN
	LBUL87_1658	
	LBUL87_1659	
	LBUL87_1660	Glycosyltransferase

LBUL87_1661	
LBUL87_1662	
LBUL87_1663	FIG00749349: hypothetical protein
LBUL87_1664	hypothetical protein
LBUL87_1665	
LBUL87_1666	
LBUL87_1667	
LBUL87_1668	Peptide methionine sulfoxide reductase MsrB
LBUL87_1669	HigA protein (antitoxin to HigB)
LBUL87_1670	
LBUL87_1671	dTDP-glucose 4,6-dehydratase
LBUL87_1672	Transcription regulator
LBUL87_1673	Glycerol uptake facilitator protein
LBUL87_1674	Phosphoenolpyruvate-dihydroxyacetone phosphotransferase, subunit DhaM; DHA-specific IIA component
LBUL87_1675	Phosphoenolpyruvate-dihydroxyacetone phosphotransferase, ADP-binding subunit DhaL
LBUL87_1676	Phosphoenolpyruvate-dihydroxyacetone phosphotransferase, dihydroxyacetone binding subunit DhaK
LBUL87_1677	
LBUL87_1678	
LBUL87_1679	
LBUL87_1680	DNA-entry nuclease (Competence-specific nuclease)
LBUL87_1681	TolA protein
LBUL87_1682	
LBUL87_1683	Hypothetical SAV0801 homolog in superantigen-encoding pathogenicity islands SaPI
LBUL87_1684	Fumarate reductase flavoprotein subunit
LBUL87_1685	
LBUL87_1686	
LBUL87_1687	Mobile element protein
LBUL87_1688	
LBUL87_1689	
LBUL87_1690	hypothetical protein
LBUL87_1691	hypothetical protein
LBUL87_1692	hypothetical protein
LBUL87_1693	
LBUL87_1694	
LBUL87_1695	capsular polysaccharide biosynthesis protein
LBUL87_1696	
LBUL87_1697	hypothetical protein
LBUL87_1698	Membrane protein involved in the export of O-antigen, teichoic acid lipoteichoic acids
LBUL87_1699	UDP-galactopyranose mutase
LBUL87_1700	
LBUL87_1701	hypothetical protein
LBUL87_1702	Glycosyltransferase
LBUL87_1703	
LBUL87_1704	glycosyltransferase
LBUL87_1705	Polysaccharide biosynthesis protein CpsF

LBUL87_1706	Undecaprenyl-phosphate galactosephosphotransferase
LBUL87_1707	capsular polysaccharide biosynthesis protein
LBUL87_1708	Tyrosine-protein kinase EpsD
LBUL87_1709	Tyrosine-protein kinase transmembrane modulator EpsC
LBUL87_1710	Exopolysaccharide biosynthesis transcriptional activator EpsA
LBUL87_1711	GTP-binding protein HflX
LBUL87_1712	L-alanyl-gamma-D-glutamyl-L-diamino acid endopeptidase
LBUL87_1713	L-alanyl-gamma-D-glutamyl-L-diamino acid endopeptidase
LBUL87_1714	Cell wall-associated hydrolase
LBUL87_1715	L-alanyl-gamma-D-glutamyl-L-diamino acid endopeptidase
LBUL87_1716	L-alanyl-gamma-D-glutamyl-L-diamino acid endopeptidase
LBUL87_1717	L-alanyl-gamma-D-glutamyl-L-diamino acid endopeptidase
LBUL87_1718	D-lactate dehydrogenase
LBUL87_1719	L-alanyl-gamma-D-glutamyl-L-diamino acid endopeptidase
LBUL87_1720	Guanylate kinase
LBUL87_1721	FIG00744133: hypothetical protein
LBUL87_1722	membrane protein, putative

Table 5.2 Characteristics of confirmed and questionable Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) arrays as predicted by CRISPRFinder in *Lb. bulgaricus* ACA-DC 87

CRISPRs	CRISPR array coordinates	CRISPR length	Direct Repeat (DR) consensus	DR length	Spacers
Confirmed	751606.. 752367	761	GTTTTAGAAGGTTGTCTATTCAATAAGGTTTAACCC	36	11
Questionable	297961.. 298038	77	AAGTGTTTTCTTATCTCATACTTT	24	1

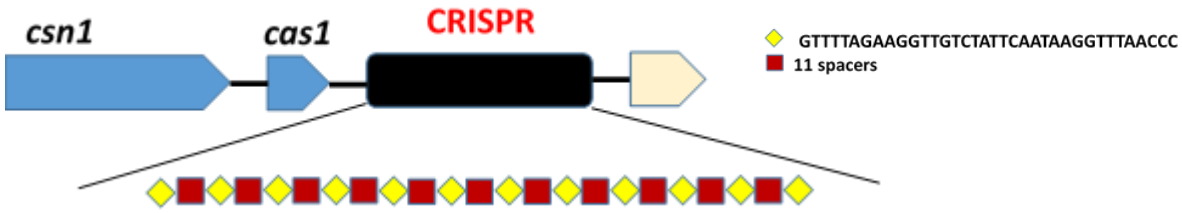


Figure 5.3. Schematic representation of the confirmed CRISPR-Cas system identified in *Lb. bulgaricus* ACA-DC 87.

Table 5.3 Blastn analysis of the spacers of the various Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated (CRISPR-Cas) systems predicted in *Lb. bulgaricus* ACA-DC 87

Coordinates of CRISPR array	Spacers	1st blastn result (excluding same species)
751606.. 752367	>spacer1 ATCCGGCAGAACTAGTTGTTCCGCCGGGAG	-
	>spacer2 TATTAACACCGACACCGACGGCCCAAACCC	-
	>spacer3 CCTCTGCTTGTTTAAGCTTGCTTATCGTTG	*Lactobacillus delbrueckii* subsp. *lactis*
	>spacer4 GGCGCCAGCAGAGCGTTATCGCGTTGCTCA	-
	>spacer5 ATTGCTTGGTTAAGCTAGCAGTTGCGTTAT	-
	>spacer6 CAATTAACTTTATTTCCTTTACGCTTTATA	-
	>spacer7 CCATAGTTAAAATCTCCTCTTTAAATTAAG	-
	>spacer8 GCAAGGCGGCCGACTCACTAGCCAGTGCTG	-
	>spacer9 CAGCCCAGGCACGGGCTCTCTTGGCTTCTG	-
	>spacer10 CATTGTTACCACGCTTTCTATAAATCCTTC	-
	>spacer11 GATCTTGTCAGAGTGGTTGAGCATGTCGTC	*Lactobacillus delbrueckii* subsp. *lactis*
297961.. 298038	>spacer1 GGATAACCTCCGTTATGTGGATTTACATGT	*Lactobacillus delbrueckii*

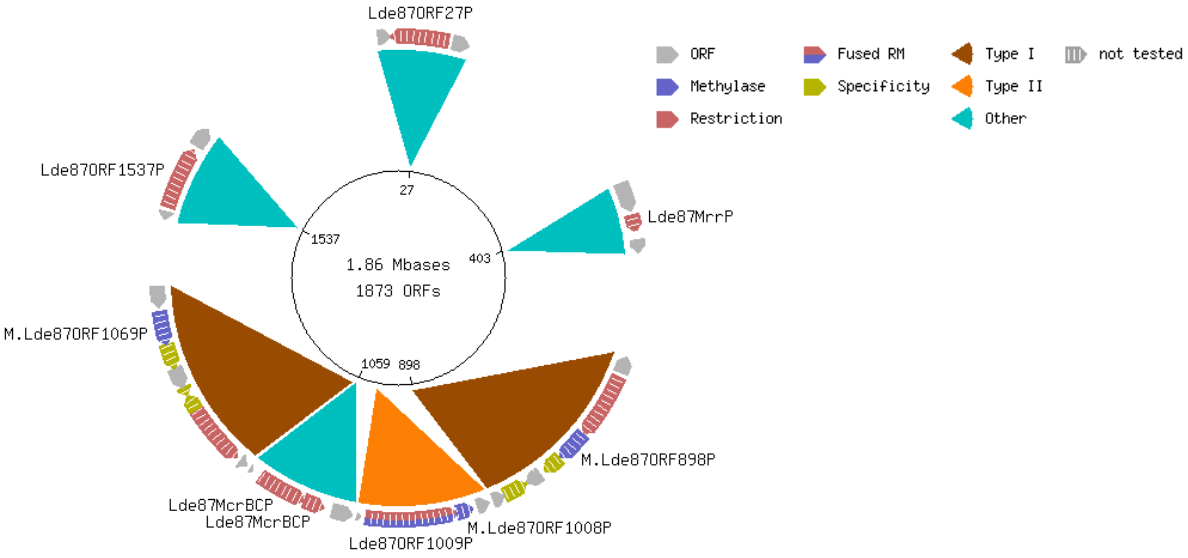
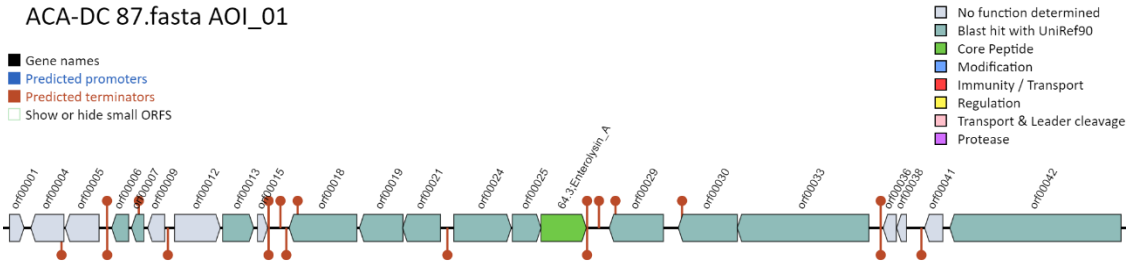
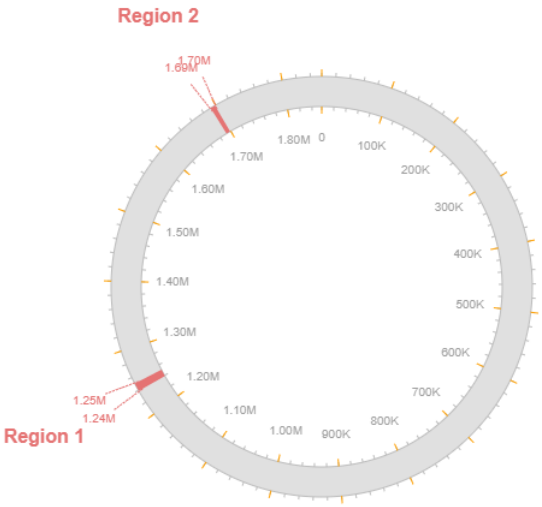


Figure 5.4 Restriction-modification systems predicted in the *Lb. bulgaricus* ACA-DC 87 chromosome by the REBASE database.



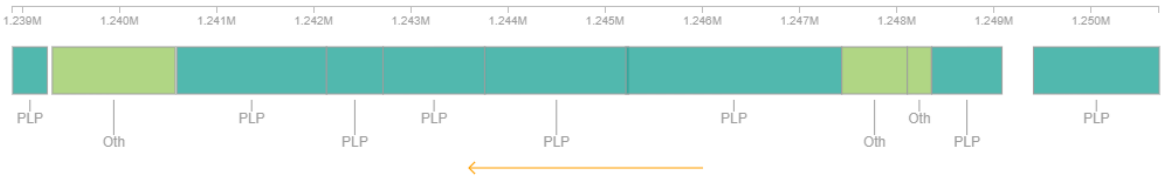
Name	Function
orf00001	
orf00004	
orf00005	
orf00006	Putative acetyltransferase SH0499 OS=Staphylococcus haemolyticus (strain JCSC1435) OX=279808 GN=SH0499 PE=3 SV=1
orf00007	Probable maltose O-acetyltransferase OS=Bacillus subtilis (strain 168) OX=224308 GN=maa PE=3 SV=1
orf00009	
orf00012	
orf00013	Signal peptidase I OS=Bacillus licheniformis OX=1402 GN=lepB PE=3 SV=1
orf00015	
orf00018	Peptidase T OS=Lactobacillus helveticus OX=1587 GN=pepT PE=1 SV=2
orf00019	GTP cyclohydrolase 1 type 2 homolog OS=Streptococcus pneumoniae serotype 4 (strain ATCC BAA-334 / TIGR4) OX=170187 GN=SP_1609 PE=1 SV=1
orf00021	tRNA (adenine(22)-N(1))-methyltransferase OS=Bacillus subtilis (strain 168) OX=224308 GN=trmK PE=1 SV=2
orf00024	Probable transposase for insertion-like sequence element IS1161 OS=Streptococcus salivarius OX=1304 PE=3 SV=1
orf00025	Uncharacterized Nudix hydrolase YvcI OS=Bacillus subtilis (strain 168) OX=224308 GN=yvcI PE=3 SV=1
64.3;Enterolysin_A	64.3;Enterolysin_A
orf00029	Uncharacterized protein YeeC OS=Bacillus subtilis (strain 168) OX=224308 GN=yeeC PE=4 SV=2
orf00030	Protein McrC OS=Escherichia coli (strain K12) OX=83333 GN=mcrC PE=1 SV=3
orf00033	5-methylcytosine-specific restriction enzyme B OS=Escherichia coli (strain K12) OX=83333 GN=mcrB PE=1 SV=3
orf00036	
orf00038	
orf00041	
orf00042	Putative type I restriction enzyme HindVIIP R protein OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) OX=71421 GN=HI_1285 PE=3 SV=1

Figure 5.5 Genes encoding for antimicrobial peptides in the genome sequence of *Lb. bulgaricus* ACA-DC 87, as predicted by BAGEL4.



Region	Region Length	Completeness	Score	# Total Proteins	Region Position	Most Common Phage	GC %	Details
1	11.8Kb	incomplete	10	11	1238903-1250705	PHAGE_Prochl_P_SSM2_NC_006883(4)	52.68%	Show
2	6.8Kb	incomplete	40	9	1691939-1698806	PHAGE_Bacill_vB_BanS_Tsamsa_NC_023007(1)	49.97%	Show

Figure 5.6 Incomplete prophage regions in the genome sequence of *Lb. bulgaricus* ACA-DC 87, as predicted by PHASTER.



#	CDS Position	BLAST Hit	E-Value
1	complement(1238903..1239259)	PHAGE_Plankt_PaV_LD_NC_016564: ABC transporter; LBUL87_1248; phage(gi371496158)	1.15e-05
2	complement(1239311..1240576)	Phosphoribosylamine--glycine ligase; LBUL87_1249	0.0
3	complement(1240592..1242133)	PHAGE_Prochl_P_SSM2_NC_006883: 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase; LBUL87_1250; phage(gi61806062)	1.09e-84
4	complement(1242134..1242715)	PHAGE_Burkho_phiE255_NC_009237: gp30, formyl transferase, putative; LBUL87_1251; phage(gi134288807)	5.83e-05
5	complement(1242715..1243764)	PHAGE_Prochl_P_SSM2_NC_006883: phosphoribosylaminoimidazole synthetase; LBUL87_1252; phage(gi61806048)	1.50e-75
6	complement(1243761..1245239)	PHAGE_Pseudo_Bjorn_NC_042103: dsDNA binding protein; LBUL87_1253; phage(gi100047)	1.34e-06
7	complement(1245215..1247437)	PHAGE_Microb_Min1_NC_009603: putative phosphoribosyl formylglycinamidine (FGAM) synthase II; LBUL87_1254; phage(gi149882852)	7.80e-172
8	complement(1247437..1248111)	Phosphoribosylformylglycinamidine synthase, glutamine amidotransferase subunit; LBUL87_1255	0.0
9	complement(1248111..1248359)	Phosphoribosylformylglycinamidine synthase, PurS subunit; LBUL87_1256	0.0
10	complement(1248364..1249086)	PHAGE_Prochl_Syn1_NC_015288: phosphoribosylaminoimidazole-succinocarboxamide synthase; LBUL87_1257; phage(gi326784002)	7.39e-48
11	complement(1249410..1250705)	PHAGE_Gordon_Ghobes_NC_031028: DNA polymerase; LBUL87_1258; phage(gi100036)	3.20e-23

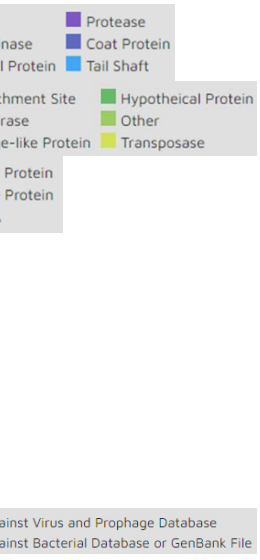
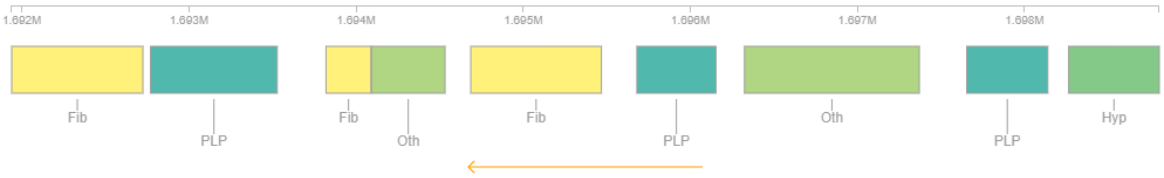


Figure 5.7 Incomplete prophage region 1 and BLAST analysis of the respective CDSs in the genome sequence of *Lb. bulgaricus* ACA-DC 87, as predicted by PHASTER.



#	CDS Position	BLAST Hit	E-Value
1	complement(1691939..1692724)	PHAGE_Strept_BillNye_NC_042105: short tail fibers; LBUL87_1712; phage(gi100205)	9.15e-11
2	complement(1692770..1693528)	PHAGE_Clostr_phiCDHM11_NC_029001: putative cell wall hydrolase protein; LBUL87_1713; phage(gi971764678)	1.58e-19
3	complement(1693820..1694089)	PHAGE_Strept_BillNye_NC_042105: short tail fibers; LBUL87_1714; phage(gi100205)	1.94e-08
4	complement(1694092..1694532)	L-alanyl-gamma-D-glutamyl-L-diamino acid endopeptidase; LBUL87_1715	0.0
5	complement(1694686..1695468)	PHAGE_Strept_BillNye_NC_042105: short tail fibers; LBUL87_1716; phage(gi100205)	3.00e-19
6	complement(1695680..1696153)	PHAGE_Clostr_phiCDHM11_NC_029001: putative cell wall hydrolase protein; LBUL87_1717; phage(gi971764678)	3.45e-18
7	complement(1696324..1697370)	D-lactate dehydrogenase; LBUL87_1718	0.0
8	complement(1697654..1698139)	PHAGE_Clostr_phiCDHM11_NC_029001: putative cell wall hydrolase protein; LBUL87_1719; phage(gi971764678)	1.14e-16
9	complement(1698264..1698806)	PHAGE_Bacill_vB_BanS_Tsamsa_NC_023007: hypothetical protein; LBUL87_1720; phage(gi564292623)	9.86e-17

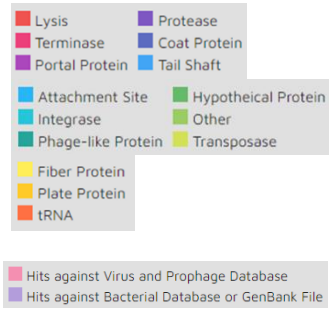


Figure 5.8 Incomplete prophage region 2 and BLAST analysis of the respective CDSs in the genome sequence of *Lb. bulgaricus* ACA-DC 87, as predicted by PHASTER.

Table 5.4 EPS gene clusters identified in *Lb. bulgaricus* ACA-DC 87 (based on BlastP similarity searches against the described EPS gene clusters of *Lb. bulgaricus* strains ATCC 11842 and 2038)

	Locus tag	RefSeq locus tag	Gene	Product
EPS cluster 1	LBUL87_1636	LBUL87_RS08555	*epsIIN, tagD*	glycerol-3-phosphate cytidylyltransferase
	LBUL87_1637	LBUL87_RS08560	*epsIIM*	LicD family protein
	LBUL87_1638	LBUL87_RS08565	*epsIIL, wzx*	oligosaccharide flippase family protein
	LBUL87_1639 (pseudogene)	LBUL87_RS08570		galactofuranosyltransferase
	LBUL87_1640	LBUL87_RS08575	*EpsIIJ*	EpsG family protein
	LBUL87_1641	LBUL87_RS08580	*EpsII-I*	DUF4422 domain-containing protein, putative glycosyltransferase
	LBUL87_1642	LBUL87_RS08585	*EpsIIH*	glycosyltransferase
	LBUL87_1643	LBUL87_RS08590	*EpsIIG*	glycosyltransferase family 2 protein
	LBUL87_1644 (pseudogene)	LBUL87_RS08595		flippase
	LBUL87_1645	LBUL87_RS08600	*EpsIIE, EpsIJ*	UDP-galactopyranose mutase
	LBUL87_1646	LBUL87_RS08605	*EpsIID*	DUF4422 domain-containing protein, putative glycosyltransferase
	LBUL87_1647	LBUL87_RS08610	*EpsIIC*	sugar transferase
	LBUL87_1648 (pseudogene)	LBUL87_RS08615		glycosyltransferase
	LBUL87_1649	LBUL87_RS08620	*EpsIIA*	LCP family protein
EPS cluster 2	LBUL87_1695	LBUL87_RS08850	*EpsIM*	glycosyltransferase
	LBUL87_1696	LBUL87_RS08855	*EpsIL* (fragment)	hypothetical protein
		LBUL87_RS08860	*EpsIL* (fragment)	hypothetical protein
	LBUL87_1698	LBUL87_RS08865	*EpsIK*	flippase
	LBUL87_1699	LBUL87_RS08870	*EpsIIE, EpsIJ*	UDP-galactopyranose mutase
	LBUL87_1700 (pseudogene)	LBUL87_RS08875		IS256 family transposase
	LBUL87_1701	LBUL87_RS08880		hypothetical protein
	LBUL87_1702	LBUL87_RS08885	*EpsIH*	DUF4422 domain-containing protein
	LBUL87_1703	LBUL87_RS08890		glycosyltransferase
	LBUL87_1704	LBUL87_RS08895	*EpsIG*	glycosyltransferase
	LBUL87_1705	LBUL87_RS08900	*epsIF*	UDP-N-acetylglucosamine--LPS N-acetylglucosamine transferase
	LBUL87_1706	LBUL87_RS08905	*EpsIE*	sugar transferase
	LBUL87_1707	LBUL87_RS08910	*EpsID, EpsD*	exopolysaccharide biosynthesis protein
	LBUL87_1708	LBUL87_RS08915	*EpsIC, EpsC*	CpsD/CapB family tyrosine-protein kinase
	LBUL87_1709	LBUL87_RS08920	*EpsIB, EpsB*	exopolysaccharide biosynthesis protein
	LBUL87_1710	LBUL87_RS08925	*EpsIA, EpsA*	LCP family protein

Table 5.5 Genes implicated in carbohydrate catabolism in *Lb. bulgaricus* ACA-DC 87

Protein	Gene	Locus tag
D-glucose transport and metabolism		
PTS system, glucose/glucoside-specific enzyme	-	LBUL87_1524, LBUL87_1562
Glucose-6-phosphate isomerase	*pgi*	LBUL87_0598
Fructose transport and metabolism		
PTS system, fructose-specific enzyme IIABC component	*fruB*	LBUL87_0162, LBUL87_1739
1-phosphofructokinase	*fruk*	LBUL87_1740
Fructose repressor	*fruR*	LBUL87_1741 pseudogene
Fructokinase	*ScrK*	LBUL87_0731
6-phosphofructokinase	*pfk2*	LBUL87_1525
Fructose-bisphosphate aldolase	*fba*	LBUL87_1350
Mannose transport and metabolism		
PTS system, mannose-specific enzyme IIABCD components	-	LBUL87_1545, LBUL87_1546, LBUL87_1547
Mannose-6-phosphate isomerase	*pmi*	LBUL87_0038
6-Phosphofructokinase	*pfK2*	LBUL87_1525

Fructose-bisphosphate aldolase	*fba*	LBUL87_1350
Lactose transport and metabolism		
PTS family lactose porter, IIABC components	*LacE/LacF*	**absence**
Transcription antiterminator	*LacT*	**absence**
6-Phospho-b-galactosidase	*LacG*	LBUL87_0246 pseudogene
Galactose-6-phosphate isomerase LacB subunit	*LacB*	LBUL87_1433 pseudogene
Galactose-6-phosphate isomerase LacA subunit	*LacA*	LBUL87_1433 pseudogene
Tagatose-6-phosphate kinase	*LacC*	**absence**
Tagatose 1,6-diphosphate aldolase	*LacD*	**absence**
Lactose permease	*LacS*	LBUL87_1042
ß-galactosidase	*LacZ*	LBUL87_1041
Lac repressor	*LacR*	LBUL87_1040 pseudogene
Saccharose transport and metabolism		
PTS family sucrose porter, EIIBC component	*ScrA*	**absence**
Sucrose 6P hydrolase (sucrase)	*ScrB*	LBUL87_1523 pseudogene
Fructokinase	*ScrK*	LBUL87_0731

Table 5.6 Genes of the proteolytic system of *Lb. bulgaricus* ACA-DC 87

Proteases and aminopeptidases		
Locus tag	**Gene**	**Protein**
LBUL87_1031	*prtB*	Proteinase precursor
LBUL87_1164		Putative metalloprotease
LBUL87_1493	*sms*	Putative ATP-dependent serine protease
LBUL87_1496	*pepC*	Aminopeptidase C
LBUL87_0432	*pepD1*	Dipeptidase
LBUL87_1761	*pepD2*	Dipeptidase
LBUL87_1732	*pepF*	Oligoendopeptidase F
LBUL87_0228	*pepG*	Aminopeptidase G
LBUL87_1599	*pepIP*	Proline iminopeptidase
LBUL87_0475	*pepM*	Methionine aminopeptidase
LBUL87_1773	*pepN*	Aminopeptidase N
LBUL87_0180	*pepO1*	Endopeptidase
LBUL87_1243	*pepP*	X-Pro dipeptidase PepP
LBUL87_1613	*pepPN*	Prolyl aminopeptidase
LBUL87_1380	*pepQ1*	X-Pro dipeptidase PepQ
LBUL87_1051	*pepT1*	Aminotripeptidase (peptidase T)
LBUL87_0312	*pepT2*	Aminotripeptidase (peptidase T)
LBUL87_1511	*pepV*	X-His dipeptidase
LBUL87_0229	*pepW*	Aminopeptidase W
LBUL87_1268	*pepX*	X-Pro dipeptidyl-peptidase
LBUL87_1474	*pepZ*	X-Pro dipeptidase PepZ

LBUL87_1541	*dpp*	Putative dipeptidyl-peptidase
LBUL87_0531	*dppA*	D-aminopeptidase
LBUL87_0509		Putative peptidase
LBUL87_0510		Putative peptidase
LBUL87_1408		Putative peptidase
LBUL87_0863		Putative acylaminoacyl-peptidase
LBUL87_1406	*gcp*	Putative glycoprotein endopeptidase
LBUL87_1019	*pcp*	Pyroglutamyl-peptidase I
LBUL87_0323	*clpC*	ATP-dependent Clp protease, ATP-binding subunit
LBUL87_0489	*clpE*	ATP-dependent Clp protease, ATP-binding subunit ClpE
LBUL87_1794	*clpL*	ATP-dependent Clp protease, ATP-binding subunit
LBUL87_0536	*clpP*	ATP-dependent Clp protease proteolytic subunit
LBUL87_0682	*clpX*	ATP-dependent Clp protease ATP-binding subunit clpX
LBUL87_1101	*hslU, clpY*	ATP-dependent protease HslVU (ClpYQ), ATPase subunit
LBUL87_1102	*hslV, clpQ*	ATP-dependent protease HslVU (ClpYQ), peptidase subunit
LBUL87_0126	*htpX1*	Protease
LBUL87_1027	*htpX2*	Protease
LBUL87_1361	*dacA*	D-alanyl-D-alanine carboxypeptidase
LBUL87_0108	*htrA*	HtrA-like serine protease
LBUL87_1341	*PrsA or PrtM*	protease maturation protein
LBUL87_1049		signal peptidase I
LBUL87_0872		signal peptidase II

Peptide and amino acid transporters

Locus tag	Gene	Protein
LBUL87_0220	*oppA3II*	Oligopeptide ABC transporter, substrate binding protein
LBUL87_0222	*oppA1II*	Oligopeptide ABC transporter, substrate binding protein
LBUL87_0223	*oppBII*	Oligopeptide ABC transporter, permease protein
LBUL87_0224	*oppCII*	Oligopeptide ABC transporter, permease protein
LBUL87_0225	*oppDII*	Oligopeptide ABC transporter, ATP-binding protein
LBUL87_0226	*oppFII*	Oligopeptide ABC transporter, ATP-binding protein
LBUL87_1202	*oppA2*	Oligopeptide ABC transporter, substrate binding protein
LBUL87_1203	*oppA1*	Oligopeptide ABC transporter, substrate binding protein
LBUL87_1204	*oppC*	Oligopeptide ABC transporter, permease protein
LBUL87_1205	*oppB*	Oligopeptide ABC transporter, permease protein
LBUL87_1206	*oppF*	Oligopeptide ABC transporter, ATP-binding protein
LBUL87_1207	*oppD*	Oligopeptide ABC transporter, ATP-binding protein
LBUL87_0311	-	Peptide binding protein
LBUL87_0688	-	Peptide binding protein
LBUL87_0973	-	Peptide binding protein
LBUL87_0438	*glnQ*	Glutamine ABC transporter, ATP-binding protein
LBUL87_0439	*glnH1*	Glutamine ABC transporter, glutamine-binding protein

LBUL87_0440	*glnM*	Glutamine ABC transporter, permease protein
LBUL87_0441	*glnP*	Glutamine ABC transporter, permease protein
LBUL87_0442	*glnH2*	Glutamine ABC transporter, glutamine-binding protein
LBUL87_1860	-	Glutamine ABC transporter, substrate binding / permease protein
LBUL87_1861	-	Glutamine ABC transporter, ATP-binding protein
LBUL87_1023	-	Glycine betaine ABC transporter, substrate binding / permease protein
LBUL87_1024	-	Glycine betaine ABC transporter, ATP-binding protein
LBUL87_0134	-	Amino Acid ABC transporter, ATP-binding protein
LBUL87_0135	-	Amino acid ABC transporter, substrate binding / permease protein
LBUL87_0205	-	Amino Acid ABC transporter, permease protein
LBUL87_0206	-	Amino acid ABC transporter, permease protein
LBUL87_0207	-	Amino acid ABC transporter, ATP-binding protein
LBUL87_0208	-	Amino acid ABC transporter, substrate binding protein
LBUL87_1128	-	Amino acid ABC transporter, ATP-binding protein
LBUL87_1129	-	Amino Acid ABC transporter, substrate binding /permease protein
LBUL87_0461	-	Amino acid permease
LBUL87_0578	-	Amino-acid permease
LBUL87_1486	-	Amino acid permease
LBUL87_1542	-	Amino acid permease
LBUL87_0411	*brnQ*	Branched-chain amino acid transport system carrier protein
LBUL87_1183	-	Glutamine-binding protein

6

Whole-genome sequence data and analysis of *Lactobacillus delbrueckii* subsp. *lactis* ACA-DC 178 isolated from Greek Kasseri cheese

Data in brief 25 (2019) 104282

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Whole-genome sequence data and analysis of *Lactobacillus delbrueckii* subsp. *lactis* ACA-DC 178 isolated from Greek Kasseri cheese

Voula Alexandraki [a], Maria Kazou [a], Bruno Pot [b], Effie Tsakalidou [a], Konstantinos Papadimitriou [a, *]

[a] *Laboratory of Dairy Research, Department of Food Science and Human Nutrition, Agricultural University of Athens, Athens, Greece*

[b] *Research Group of Industrial Microbiology and Food Biotechnology (IMDO), Department of Bioengineering Sciences (DBIT), Vrije Universiteit Brussel, Brussels, Belgium*

ARTICLE INFO

Article history:
Received 24 June 2019
Accepted 10 July 2019
Available online 16 July 2019

Keywords:
Lactic acid bacteria
Dairy
Fermentation
Genomics
CRISPR
Genomic island
Milk

ABSTRACT

Lactobacillus delbrueckii subsp. *lactis* is employed in the production of various types of cheese. Here, we report the complete genome sequence of *L. lactis* ACA-DC 178 isolated from Greek Kasseri cheese. The chromosome of ACA-DC 178 contains 2,050,316 bp with a GC content of 49.6%. A total of 2,112 genes were identified in the genome sequence including 1,752 protein-coding genes, 239 putative pseudogenes, 94 tRNA and 27 rRNA genes. According to the COG annotation, about 80% of the protein-coding genes (1,417 proteins) were assigned to at least one functional category. Approximately the 1/3 of these proteins were distributed among three categories, namely replication, recombination and repair (category L: 10.6%), translation, ribosomal structure and biogenesis (category J: 7.5%) and amino acid transport and metabolism (category E: 7.2%). Fourteen integrated GIs with a total of 159 genes were found in ACA-DC 178 genome. Several of these genes encode proteins associated with exopolysaccharide biosynthesis, amino acid transport and subunits of restriction-modification systems. One large CRISPR array of 3,197 bp containing 52 spacers, several of which are identical to phage sequences having hosts in the genus *Lactobacillus*, was also identified. The annotated genome sequence of *L. lactis* ACA-DC 178 is deposited at the European

* Corresponding author.
E-mail address: kpapadimitriou@aua.gr (K. Papadimitriou).

https://doi.org/10.1016/j.dib.2019.104282

Nucleotide Archive under the accession number LS991409. Raw reads are deposited in the Sequence Read Archive (SRR8866601-3).

Specifications Table

Subject	Microbiology
Specific subject area	Genome analysis
Type of data	Table, Figure
How data were acquired	Genome sequencing: Illumina HiSeq 2000 (Illumina, CA), Denovo sequence assembly: SOAPdenovo v.2.04 software, MapSolver software (OpGen Technologies, Inc., Madison, WI), Bioinformatics approaches: Rapid Annotation using Subsystem Technology (RAST) v.2.0, Prodigal, MetaGeneAnnotator, FGENESB, GenePRIMP pipeline, eggNOG-mapper v.4.5, IslandViewer 4, CRISPRFinder
Data format	Raw sequence reads and genome assembly and annotation
Parameters for data collection	Genomic DNA from pure culture
Description of data collection	Purification of genomic DNA, genome sequencing, genome assembly and annotation
Data source location	Traditional Greek Kasseri Cheese, Athens, Greece
Data accessibility	Data are deposited in the respective databases and are publicly available. The annotated whole-genome sequence of *L. lactis* ACA-DC 178 is deposited at the European Nucleotide Archive (ENA) under the accession number LS991409 (https://www.ebi.ac.uk/ena/data/view/LS991409). Raw sequence reads are deposited in the Sequence Read Archive (SRA; SRR8866601-3; https://www.ncbi.nlm.nih.gov/sra/?term=SRR8866601; https://www.ncbi.nlm.nih.gov/sra/?term=SRR8866602; https://www.ncbi.nlm.nih.gov/sra/?term=SRR8866603).

Value of the Data

- *Lactobacillus delbrueckii* subsp. *lactis* is an important bacterium used in cheese production. For this reason analysis of the genome sequence of strain ACA-DC 178 will provide valuable information for its adaptation in the milk environment and its technological properties.
- Data presented here can be valuable for researchers involved in the field of genomic analysis of lactic acid bacteria and food fermentations.
- Data can be used by researchers to perform comparative and functional genomics to further shed light in the evolution, biology and technological properties of the *L. delbrueckii* species.
- Increasing the number of complete genome sequences within *L. delbrueckii* will further aid our understanding of this species.

1. Data

In this study, we present the complete genome sequence of *L. lactis* ACA-DC 178 isolated from Greek Kasseri cheese [1,2]. The *L. delbrueckii* species consists of six subspecies, including *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Lactobacillus delbrueckii* subsp. *lactis*, *Lactobacillus delbrueckii* subsp. *delbrueckii*, *Lactobacillus delbrueckii* subsp. *indicus*, *Lactobacillus delbrueckii* subsp. *sunkii* and *Lactobacillus delbrueckii* subsp. *jakobsenii* [3,4]. *L. lactis* is the second subspecies used as a starter in the dairy industry along with *L. bulgaricus* within the *L. delbrueckii* species [3]. The *in silico* assembly of the ACA-DC 178 chromosome was validated against a *Nhe*I whole-genome optical map of the strain (Fig. 1). Our assembly presented 100% matching between the *Nhe*I restriction sites of the optical map and the relevant sites in our genome sequence *in silico* digested with the same enzyme. The genome was found to be 2,050,316 bp with a GC content of 49.6%. We were able to annotate a total of 2,112 genes, including

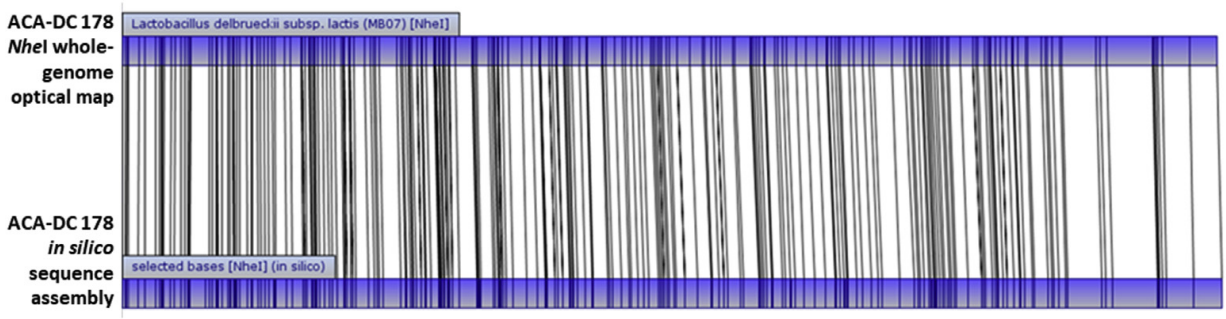


Fig. 1. Validation of the *L. lactis* ACA-DC 178 genome assembly. Alignment of the *in silico* genome assembly of *L. lactis* ACA-DC 178 (bottom) against a *Nhe*I whole-genome optical map of the strain (top).

1,752 protein-coding genes, 239 putative pseudogenes, 94 tRNA and 27 rRNA genes (Fig. 2). Further analysis revealed that about 80% of the protein-coding genes (1,417 proteins) could be assigned to at least one Cluster of Orthologous Groups (COG) functional category. Most of these proteins (approximately 1/3) were distributed among three categories related to housekeeping processes, namely replication, recombination and repair (category L: 10.6%), translation, ribosomal structure and biogenesis (category J: 7.5%) and amino acid transport and metabolism (category E: 7.2%) (Table 1). Additional features of the ACA-DC 178 included 14 integrated genomic islands (GIs; Fig. 3) and a clustered regularly interspaced short palindromic repeats-CRISPR-associated (CRISPR-Cas) system (Fig. 4). The GIs carry 159 genes some of which could be assigned to functions like exopolysaccharide biosynthesis, amino acid transport and restriction-modification. The CRISPR array was relatively long, consisting of 3,197 bp and 52 spacers. Detailed analysis of the spacers identified several segments of phage sequences, which have hosts belonging to the *Lactobacillus* genus.

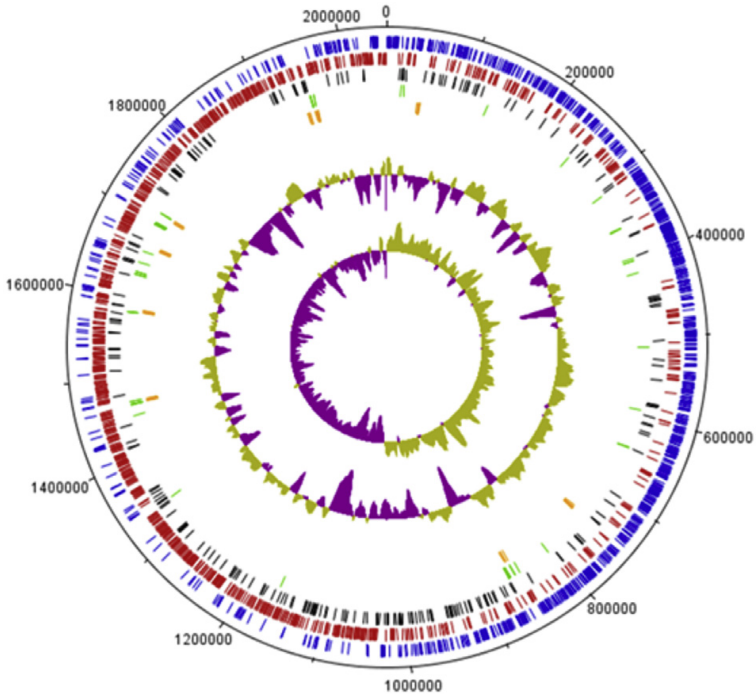


Fig. 2. Circular map of the *L. lactis* ACA-DC 178 genome. Each ring represents specific genomic features appearing from the periphery to the centre of the map: Forward CDSs (blue); Reverse CDSs (red); Pseudogenes (black); tRNA (green); rRNA (orange); %GC plot; GC skew.

Table 1
Distribution of *L. lactis* ACA-DC 178 proteins in COG categories.

	COG	Proteins	%	Description
Information storage and processing	J	132	7.5	Translation, ribosomal structure and biogenesis
	K	100	5.7	Transcription
	L	186	10.6	Replication, recombination and repair
Cellular processes and signaling	D	19	1.1	Cell cycle control, cell division, chromosome partitioning
	M	91	5.2	Cell wall/membrane biogenesis
	N	5	0.3	Cell motility
	O	47	2.7	Posttranslational modification, protein turnover, chaperones
	T	52	3.0	Signal transduction mechanisms
	U	19	1.1	Intracellular trafficking and secretion
	V	44	2.5	Defense mechanisms
Metabolism	C	46	2.6	Energy production and conversion
	E	127	7.2	Amino acid transport and metabolism
	F	67	3.8	Nucleotide transport and metabolism
	G	102	5.8	Carbohydrate transport and metabolism
	H	30	1.7	Coenzyme transport and metabolism
	I	32	1.8	Lipid transport and metabolism
	P	73	4.2	Inorganic ion transport and metabolism
	Q	3	0.2	Secondary metabolites biosynthesis, transport and catabolism
Poorly characterized	S	262	15.0	Function unknown
–	–	335	19.1	Not in COGs

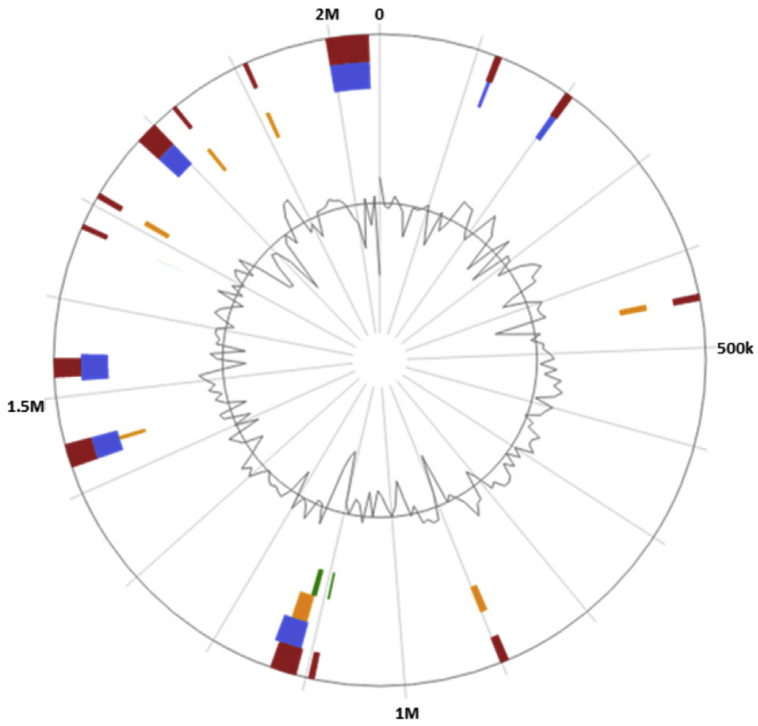


Fig. 3. Circular map of the *L. lactis* ACA-DC 178 genome. Highlighted regions correspond to GIs. GIs are colored within the circular map according to the prediction method used: green, orange and blue were predicted by IslandPick, SIGI-HMM and IslandPath-DIMOB, respectively. The integrated GIs are presented on the periphery in red. The black line plot represents the GC content (%) of the genomic sequence.

2. Experimental design, materials, and methods

L. lactis ACA-DC 178 was grown overnight in MRS broth (Merck, Darmstadt, Germany) at 30 °C. DNA was extracted according to a previously published protocol [5]. The genome was sequenced at the Beijing Genomics Institute (BGI Co., Ltd, Hong Kong) using the Illumina HiSeq 2000 platform (Illumina,

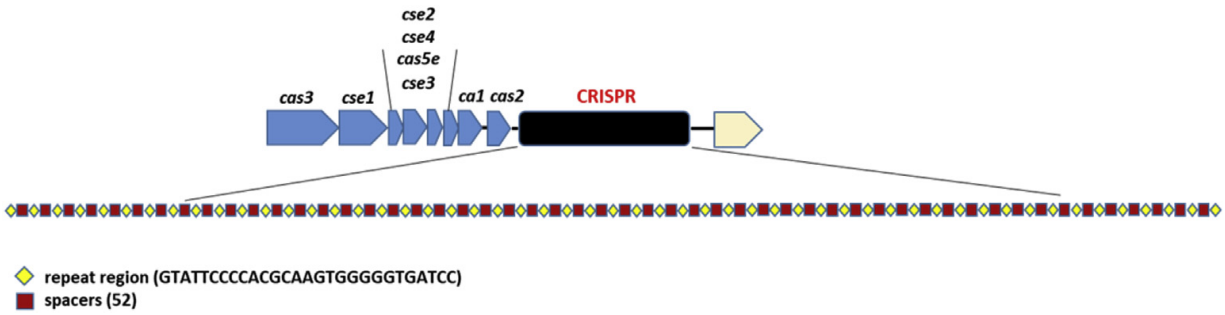


Fig. 4. Graphical presentation of the CRISPR-Cas system of *L. lactis* ACA-DC 178.

CA) employing paired-end libraries of 500 bp, 2,000 bp and 6,000 bp. The assembly of reads with SOAPdenovo v.2.04 [6] resulted in one circular chromosome that was verified against a *Nhe*I whole-genome optical map of the strain [7] produced at Microbion SRL (Verona, Italy). The alignment between the assembly and the optical map was performed with the MapSolver software (OpGen Technologies, Inc., Madison, WI). The ACA-DC 178 genome sequence was analyzed using Prodigal [8], MetaGeneAnnotator [9] and FGENESB [10] gene prediction programs. Genome annotation and prediction of rRNA and tRNA genes was performed with RAST v.2.0 [11] and putative pseudogenes were predicted with the GenePRIMP pipeline [12]. The results of the analysis were optimized with manual curation. COG annotations were computed using eggNOG-mapper based on eggNOG v.4.5 orthology database [13]. Further bioinformatic analysis was performed for the identification of GIs with IslandViewer 4 [14] and CRISPR with CRISPRFinder [15].

Acknowledgments

We would like to thank Dr. Nikos Kyrpides at the Joint Genome Institute (United States Department of Energy) for the analysis of the ACA-DC 178 genome with the Gene PRIMP pipeline. The present work was co-financed by the European Social Fund and the National resources EPEAEK and YPEPTH through the Thales project.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E. Tsakalidou, R. Anastasiou, I. Vandenberghe, J. Beeumen, G. Kalantzopoulos, Cell wall bound proteinase of *Lactobacillus delbrueckii* subsp. *lactis* ACA-DC 178. Characterisation and specificity towards β-casein, Appl. Environ. Microbiol. 65 (1999) 2035–2040.

[2] E. Tsakalidou, E. Manolopoulou, E. Kabaraki, E. Zoidou, B. Pot, K. Kersters, G. Kalantzopoulos, The combined use of whole cell protein extracts for the identification (SDS-PAGE) and enzyme activity screening of lactic acid bacteria isolated from traditional Greek dairy products, System. Appl. Microbiol. 17 (1994) 444–458. https://doi.org/10.1016/S0723-2020(11)80062-7.

[3] H. El Kafsi, J. Binesse, V. Loux, J. Buratti, S. Boudebbouze, R. Dervyn, S. Kennedy, N. Galleron, B. Quinquis, J.M. Batto, B. Moumen, E. Maguin, M. van de Guchte, *Lactobacillus delbrueckii* ssp. *lactis* and ssp. *bulgaricus*: a chronicle of evolution in action, BMC Genomics 15 (1) (2014) 407. https://doi.org/10.1186/1471-2164-15-407.

[4] D.B. Adimpong, D.S. Nielsen, K.I. Sørensen, F.K. Vogensen, H. Sawadogo-Lingani, P.M. Derkx, L. Jespersen, *Lactobacillus delbrueckii* subsp. *jakobsenii* subsp. nov., isolated from dolo wort, an alcoholic fermented beverage in Burkina Faso, Int. J. Syst. Evol. Microbiol. 63 (2013) 3720–3726. https://doi.org/10.1099/ijs.0.048769-0.

[5] D.G. Pitcher, N.A. Saunders, R.J. Owen, Rapid extraction of bacterial genomic DNA with guanidium thiocyanate, Lett. Appl. Microbiol. 8 (1989) 151–156. https://doi.org/10.1111/j.1472-765X.1989.tb00262.x.

[6] R. Li, Y. Li, K. Kristiansen, J. Wang, SOAP: short oligonucleotide alignment program, Bioinformatics 24 (2008) 713–714. https://doi.org/10.1093/bioinformatics/btn025.

[7] P. Latreille, S. Norton, B.S. Goldman, J. Henkhaus, N. Miller, B. Barbazuk, H.B. Bode, C. Darby, Z. Du, S. Forst, S. Gaudriault, B. Goodner, H. Goodrich-Blair, S. Slater, Optical mapping as a routine tool for bacterial genome sequence finishing, BMC Genomics 8 (2007) 321. https://doi.org/10.1186/1471-2164-8-321.

[8] D. Hyat, G.L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification, BMC Bioinf. 11 (2010) 119. https://doi.org/10.1186/1471-2105-11-119.
[9] H. Noguchi, T. Taniguchi, T. Itoh, MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes, DNA Res. 15 (2008) 387–396. https://doi.org/10.1093/dnares/dsn027.
[10] V. Solovyev, A. Salamov, Automatic annotation of microbial genomes and metagenomic sequences, in: R.W. Li (Ed.), Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies, Nova Science Publishers, New York, 2011, pp. 61–78.
[11] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server: rapid annotations using subsystems technology, BMC Genomics 9 (2008) 75. https://doi.org/10.1186/1471-2164-9-75.
[12] A. Pati, N.N. Ivanova, N. Mikhailova, G. Ovchinnikova, S.D. Hooper, A. Lykidis, N.C. Kyrpides, GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes, Nat. Methods 7 (2010) 455–457. https://doi.org/10.1038/nmeth.1457.
[13] J. Huerta-Cepas, K. Forslund, L.P. Coelho, D. Szklarczyk, L.J. Jensen, C. von Mering, P. Bork, Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper, Mol. Biol. Evol. 34 (8) (2017) 2115–2122. https://doi.org/10.1093/molbev/msx148.
[14] C. Bertelli, M.R. Laird, K.P. Williams, Simon Fraser University Research Computing Group, B.Y. Lau, G. Hoad, G.L. Winsor, F.S. L. Brinkman, IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets, Nucleic Acids Res. 45 (2017) W30–W35. https://doi.org/10.1093/nar/gkx343.
[15] I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats, Nucleic Acids Res. 35 (2007) W52–W57. https://doi.org/10.1093/nar/gkm360.

Supporting information

Table 6.1 Genes in the integrated genomic islands identified in the genome of *Lb. lactis* ACA-DC 178

Genomic islands	Locus_tag	Product
GI_1	ACADC178_0122	ATPase component BioM of energizing module of biotin ECF transporter
	ACADC178_0123	hypothetical protein
	ACADC178_0124	hypothetical protein
	ACADC178_0125	hypothetical protein
	ACADC178_0126	SOS-response repressor and protease LexA (EC 3.4.21.88)
	ACADC178_0127	Mobile element protein
	ACADC178_0128	FIG00743963: hypothetical protein
GI_2	ACADC178_0199	transposase
	ACADC178_0200	Mobile element protein
	ACADC178_0201	
	ACADC178_0202	hypothetical protein
	ACADC178_0203	hypothetical protein
	ACADC178_0204	Methionine ABC transporter ATP-binding protein
	ACADC178_0205	hypothetical protein
	ACADC178_0206	DNA-binding response regulator
	ACADC178_0207	hypothetical protein
	ACADC178_0208	Purine trans deoxyribosylase (Nucleoside deoxyribosyltransferase-I)
	ACADC178_0209	Phosphotransferase system galacitol-specific IIA domain (Ntr-type)
	ACADC178_0210	transposase
GI_3	ACADC178_0450	Positive transcriptional regulator, MutR family
	ACADC178_0451	Siderophore biosynthesis L-2,4-diaminobutyrate decarboxylase
	ACADC178_0452	Peptide synthetase
	ACADC178_0453	
GI_4	ACADC178_0882	
	ACADC178_0883	Dihydroneopterin aldolase (EC 4.1.2.25)
	ACADC178_0884	2-amino-4-hydroxy-6- hydroxymethyldihydropteridine pyrophosphokinase / GTP cyclohydrolase I type 1
	ACADC178_0885	Dihydrofolate synthase (EC 6.3.2.12) @ Folylpolyglutamate synthase
	ACADC178_0886	Dihydropteroate synthase
	ACADC178_0887	Adenosylhomocysteinase (EC 3.3.1.1)
	ACADC178_0888	
	ACADC178_0889	
	ACADC178_0890	hypothetical protein
	ACADC178_0891	Transcriptional regulator, MerR family
	ACADC178_0892	polar amino acid ABC transporter permease
	ACADC178_0893	ABC transporter

	ACADC178_0894	ABC transporter ATPase component
GI_5	ACADC178_1076	Galactose operon repressor, GalR-LacI family of transcriptional regulators
	ACADC178_1077	Beta-galactosidase (EC 3.2.1.23)
	ACADC178_1078	Lactose and galactose permease, GPH translocator family
	ACADC178_1079	
GI_6	ACADC178_1093	
	ACADC178_1094	
	ACADC178_1095	hypothetical protein
	ACADC178_1096	hypothetical protein
	ACADC178_1097	
	ACADC178_1098	
	ACADC178_1099	
	ACADC178_1100	Type I restriction-modification system, DNA-methyltransferase subunit M
	ACADC178_1101	
	ACADC178_1102	Restriction endonuclease S subunits
	ACADC178_1103	
	ACADC178_1104	
	ACADC178_1105	hypothetical protein
	ACADC178_1106	hypothetical protein
	ACADC178_1107	
	ACADC178_1108	
	ACADC178_1109	Mobile element protein
	ACADC178_1110	FIG00754632: hypothetical protein
	ACADC178_1111	possible DNA helicase
	ACADC178_1112	Mobile element protein
	ACADC178_1113	FIG00744915: hypothetical protein
	ACADC178_1114	FIG00748645: hypothetical protein
GI_7	ACADC178_1394	Foldase protein PrsA precursor (EC 5.2.1.8) @ Foldase clustered with pyrimidine conversion
	ACADC178_1395	5-Enolpyruvylshikimate-3-phosphate synthase (EC 2.5.1.19)
	ACADC178_1396	Shikimate kinase I (EC 2.7.1.71)
	ACADC178_1397	3'->5' exoribonuclease Bsu YhaM
	ACADC178_1398	DNA double-strand break repair Rad50 ATPase
	ACADC178_1399	DNA repair exonuclease family protein YhaO
	ACADC178_1400	Hypothetical protein SAV1845
	ACADC178_1401	Multimodular transpeptidase-transglycosylase
	ACADC178_1402	Similar to ribosomal large subunit pseudouridine synthase D, Bacillus subtilis YhcT type
	ACADC178_1403	Fructose-bisphosphate aldolase class II (EC 4.1.2.13)
	ACADC178_1404	D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4)
	ACADC178_1405	
	ACADC178_1406	FIG00753989: hypothetical protein

	ACADC178_1407	Mobile element protein
	ACADC178_1408	FIG00751841: hypothetical protein
	ACADC178_1409	D-alanyl-D-alanine carboxypeptidase (EC 3.4.16.4)
	ACADC178_1410	NAD(P)HX dehydratase
	ACADC178_1411	D-alanyl-D-alanine carboxypeptidase
	ACADC178_1412	Pyruvate oxidase (EC 1.2.3.3)
GI_8	ACADC178_1473	FIG00751417: hypothetical protein
	ACADC178_1474	FIG00745302: hypothetical protein
	ACADC178_1475	
	ACADC178_1476	hypothetical protein
	ACADC178_1477	hypothetical protein
	ACADC178_1478	Beta-hexosaminidase (EC 3.2.1.52)
	ACADC178_1479	FmtA protein involved in methicillin resistance / affects cell wall cross-linking and amidation
	ACADC178_1480	
	ACADC178_1481	hypothetical protein
	ACADC178_1482	Osmotically activated L-carnitine/choline ABC transporter, permease protein OpuCD
	ACADC178_1483	Osmotically activated L-carnitine/choline ABC transporter, substrate-binding protein OpuCC
	ACADC178_1484	Osmotically activated L-carnitine/choline ABC transporter, permease protein OpuCB
	ACADC178_1485	Osmotically activated L-carnitine/choline ABC transporter, ATP-binding protein OpuCA
	ACADC178_1486	FIG00749320: hypothetical protein
	ACADC178_1487	DNA-binding protein HBsu
	ACADC178_1488	
	ACADC178_1489	
	ACADC178_1490	LSU ribosomal protein L7/L12 (P1/P2)
	ACADC178_1491	LSU ribosomal protein L10p (P0)
	ACADC178_1492	Mobile element protein
	ACADC178_1493	Hypothetical membrane protein
GI_9	ACADC178_1621	
	ACADC178_1622	Putative iron-sulfur cluster assembly scaffold protein for SUF system, SufE2
	ACADC178_1623	Cysteine desulfurase (EC 2.8.1.7)
	ACADC178_1624	Iron-sulfur cluster assembly protein SufD
	ACADC178_1625	Iron-sulfur cluster assembly ATPase protein SufC
GI_10	ACADC178_1652	hypothetical protein
	ACADC178_1653	hypothetical protein
GI_11	ACADC178_1722	hypothetical protein
	ACADC178_1723	putative glycosyltransferase

	ACADC178_1724	Membrane protein involved in the export of O-antigen, teichoic acid lipoteichoic acids
	ACADC178_1725	UDP-galactopyranose mutase (EC 5.4.99.9)
	ACADC178_1726	Mobile element protein
	ACADC178_1727	Cell envelope-associated transcriptional attenuator LytR-CpsA-Psr, subfamily F2 (as in PMID19099556)
	ACADC178_1728	hypothetical protein
	ACADC178_1729	Cell envelope-associated transcriptional attenuator LytR-CpsA-Psr, subfamily F2 (as in PMID19099556)
	ACADC178_1730	Mobile element protein
	ACADC178_1731	
	ACADC178_1732	DNA-entry nuclease (Competence-specific nuclease) (EC 3.1.30.-)
	ACADC178_1733	hypothetical protein
	ACADC178_1734	hypothetical protein
	ACADC178_1735	Transcriptional regulator, XRE family
	ACADC178_1736	Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94)
	ACADC178_1737	
	ACADC178_1738	Mobile element protein
	ACADC178_1739	hypothetical protein
	ACADC178_1740	hypothetical protein
	ACADC178_1741	hypothetical protein
	ACADC178_1742	FIG007491: hypothetical protein YeeN
	ACADC178_1743	
	ACADC178_1744	FIG00749349: hypothetical protein
	ACADC178_1768	hypothetical protein
	ACADC178_1769	Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase (EC 2.4.1.-)
GI_12	ACADC178_1770	Glycosyltransferase
	ACADC178_1771	galactoside O-acetyltransferase
	ACADC178_1772	Glycosyltransferase
	ACADC178_1852	
	ACADC178_1853	Lead, cadmium, zinc and mercury transporting ATPase (EC 3.6.3.3) (EC 3.6.3.5); Copper-translocating
GI_13	ACADC178_1854	
	ACADC178_1855	
	ACADC178_1856	
	ACADC178_1932	Mobile element protein
	ACADC178_1933	
	ACADC178_1934	FIG00752231: hypothetical protein
GI_14	ACADC178_1935	hypothetical protein
	ACADC178_1936	FIG00742910: hypothetical protein
	ACADC178_1937	General stress protein, Gls24 family
	ACADC178_1938	hypothetical protein

ACADC178_1939	hypothetical protein
ACADC178_1940	Mobile element protein
ACADC178_1941	Mobile element protein
ACADC178_1942	
ACADC178_1943	Deoxyadenosine kinase / Deoxyguanosine kinase
ACADC178_1944	Deoxyadenosine kinase (EC 2.7.1.76) / Deoxyguanosine kinase
ACADC178_1945	/ Glutamine transport system permease protein GlnP (TC 3.A.1.3.2)
ACADC178_1946	Glutamate transport ATP-binding protein
ACADC178_1947	Xanthine/uracil/thiamine/ascorbate permease family protein
ACADC178_1948	transport protein (putative)
ACADC178_1949	Molybdopterin biosynthesis protein MoeB
ACADC178_1950	hypothetical protein
ACADC178_1951	Mobile element protein
ACADC178_1952	Mobile element protein
ACADC178_1953	hypothetical protein
ACADC178_1954	Mobile element protein
ACADC178_1955	Positive transcriptional regulator, MutR family
ACADC178_1956	ATP-dependent DNA helicase
ACADC178_1957	hypothetical protein
ACADC178_1958	hypothetical protein
ACADC178_1959	Voltage-gated chloride channel family protein
ACADC178_1960	
ACADC178_1961	
ACADC178_1962	Voltage-gated chloride channel family protein
ACADC178_1963	Three-component quorum-sensing regulatory system, response regulator
ACADC178_1964	Accessory gene regulator C (sensor histidine kinase)
ACADC178_1965	hypothetical protein
ACADC178_1966	hypothetical protein
ACADC178_1967	hypothetical protein
ACADC178_1968	Lanthionine biosynthesis protein LanL
ACADC178_1969	Multidrug resistance ABC transporter ATP-binding and permease protein
ACADC178_1970	hypothetical protein
ACADC178_1971	extracellular zinc metalloproteinase
ACADC178_1972	FIG00749399: hypothetical protein
ACADC178_1973	FIG00753649: hypothetical protein
ACADC178_1974	hypothetical protein
ACADC178_1975	FIG00742395: hypothetical protein
ACADC178_1976	FIG00742190: hypothetical protein
ACADC178_1977	ABC transporter, ATP-binding protein
ACADC178_1978	FIG00747667: hypothetical protein
ACADC178_1979	hypothetical protein
ACADC178_1980	FIG00748020: hypothetical protein

ACADC178_1981 FIG00742395: hypothetical protein

Table 6.2 Characteristics of confirmed and questionable Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) arrays as predicted by CRISPRFinder in *Lb. lactis* ACA-DC 178

CRISPRs	CRISPR array coordinates	CRISPR length	Direct Repeat (DR) consensus	DR length	Spacers
Confirmed	818084..821281	3,197	GTATTCCCCACGCAAGTGGGGGTGATCC	28	52
Questionable	306504..306598	94	AGGCCAGCTCTAAGTCCGCGGCT	23	1
Questionable	1207296..1207399	103	CCGCTGCCCCCGCACTTGTCACAGGT	26	1

Table 6.3 Blastn analysis of the spacers of the various Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated (CRISPR-Cas) systems predicted in *Lb. lactis* ACA-DC 178

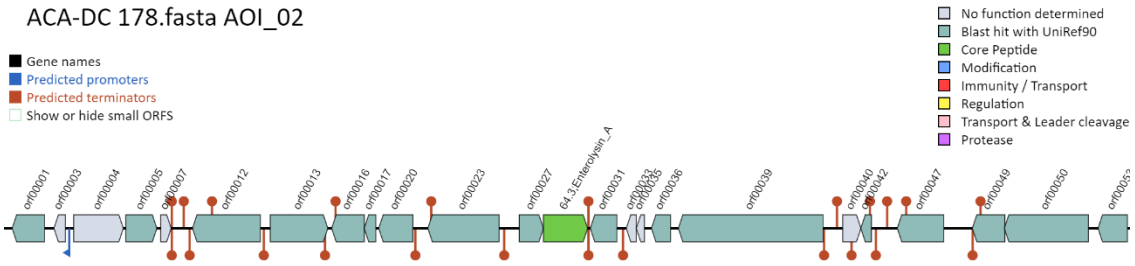
Coordinates of CRISPR array	Spacers	1st blastn result (excluding same species)
818084..821281	>spacer1 TGTTGATTTCTCAAGGCTCTTTCTTTCGGTTAT	*Lactobacillus* phages ViSo-2018b/phiJB
	>spacer2 TATGTTATTTAACCAAAAGGAATTAACCGAACA	*Lactobacillus* phages Ld3/Ld17/c5
	>spacer3 TCATTAGCATTAGCTGGGGTTTGTGGATCTTGT	-
	>spacer4 TATTTGTTTTTACCACCCGGCCAAGTAAAGGTG	*Lactobacillus* phage LfeInf
	>spacer5 TTGTGGATACCCCTATTAACAAAAAAGACGAGGG	*Lactobacillus* phage P1174
	>spacer6 TTTTGGCCAACAGGGGCAGTTCACGGCACTAACA	*Lactobacillus* phage LL-H
	>spacer7 TACACGTGGTCTAGACTCAAGGGGGATGCTGGT	*Lactobacillus* phage phiJB
	>spacer8 TCATCAAGTGCGCCGCCATCAAAGCCCTTCGTG	*Lactobacillus* phages ViSo-2018b/phiJB
	>spacer9 CTGATAACATCCCCGCTGGTGGTAAGCTTAGCTT	*Lactobacillus* phage Ld25A
	>spacer10 TATCGGTGTCTTGACTGACTACATGCAGGCAGA	*Lactobacillus* phage phiJB
	>spacer11 AGTGTGTCACCCATCAAGGTAGGCCATGCACGG	-
	>spacer12 TATCGAAAATACACGGCTATCTGAGGAGCTTGA	*Lactobacillus* phage phiJB
	>spacer13 CCGTGTTGATTAACGGGGAACCGTGGTTTGTTG	*Escherichia coli* plasmids AH01/pAMSC3/ p634-1
	>spacer14 CGCAGCGGCAGAAGAATACGTACACCTTTACTT	-
	>spacer15 TGCAGTCTACACCTGGACGGGCACAACCTGGGA	-
	>spacer16 TAGCGATAACGCAACGTTGGAAACCCAATGTTG	-
	>spacer17 AAGCACCCAATAGCAGAGTATATTAAAATCCCA	*Lb. delbrueckii* subsp. *delbrueckii*, *Lb. johnsonii*
	>spacer18 TCGTCGCCCGGTACTCGCCCAATTTCAAACTGC	*Lactobacillus* phage ViSo-2018b
	>spacer19 TTACAGCATTACTGACATTTTTATACAAAGGAG	-

Table 6.3 (continued)

Coordinates of CRISPR array	Spacers	1st blastn result (excluding same species)
818084..821281	>spacer20 TGTTTTCAATCCATTTTGTCATTGTTTTCCTCC	*Lactobacillus* phage phiJB
	>spacer21 TTCTGGTTTGGCGTCACTATCTCCGACATCGA	-
	>spacer22 TCAACTTACGTCAGTTCTGACGGGATTACCTCA	-
	>spacer23 CAATGCTCTCTATGTCGGGACTTATACAGACTA	-
	>spacer24 TATATAAATGCTAATTTCCCATAAATTACTACT	*Lb. delbrueckii* subsp. *bulgaricus*
	>spacer25 CATGTACGACGTTGAAAACAAAATGGTCGCAGA	-
	>spacer26 ATGGTGCTCGTCCAAGTAACCCAGGTAGGCGA	-
	>spacer27 TGTTGTCATTGGTGACATTGCTGAGGAATGGA	-
	>spacer28 TCAATGCAGAGCAGGCTGAGACGTTCGACAAGG	-
	>spacer29 CGAAACTGACACAACTGCTGGTGCATATGTTGG	-
	>spacer30 CCCAAAAGTGAGACCCTTTTGCCAGCCATTTTT	-
	>spacer31 TGCGGGGCGCTTAGAGCGCATGGTGCGAGCGAG	-
	>spacer32 CCCAGCTAAAAGACCAATCCCGCCAGTCAAAAG	*Lactobacillus* phage phiJB
	>spacer33 TTGGCAAGCTCCACAAGCGGCCGATTTTGGCTCA	-
	>spacer34 TTGCACCGTAGGCGGTCATGAATCCGCTAGCTA	-
	>spacer35 TTGTCTTTTTGGCAGCCCGCTTGGTCGTCTTCT	-
	>spacer36 TTCCAGCGGACGCCACAATAGCTCCGATGCCAA	-
	>spacer37 CTCAATTGCCAACGTCATCAACTCTGTCGGTAA	-
	>spacer38 TATGATTATGCAAGGGGGTCTGGCTAATGTGG	-
	>spacer39 CTGCCCGTCTGCTGTATAGACATAAGTTTGCCG	-
	>spacer40 CGATCAGACCAAAGCGACTCTAAAGCGAATCAA	-
	>spacer41 TTAAATTAAAAGAGGTGTTAAAAATGACTATTC	-
	>spacer42 CGCCCAAAAGCTTTTTGCGCCCCCCGTCTTTC	-
	>spacer43 TCAGTAGAAGAGATATTGAGGTAAAATGTACTG	-
	>spacer44 TCGTTTCTGGTGTTGCGTACGCCCCGTGCTTGC	*Lactobacillus* phage LL-H
	>spacer45 CGCGTTTTTCCGAGCAGGCATCAAGGACCAAG	-
	>spacer46 TAGTCTATAGGGGATCCCCGGCTTTTTGGATAA	-

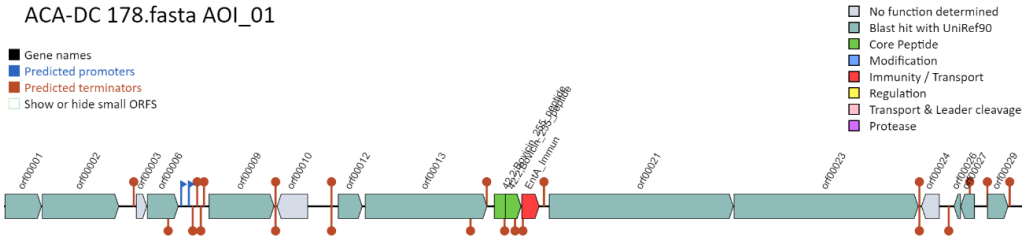
Table 6.3 (continued)

Coordinates of CRISPR array	Spacers	1st blastn result (excluding same species)
818084..821281	>spacer47 CATCGAGTCGGCCACACCGCCGATCGCAGACGT	-
	>spacer48 TTCAAGTGCTTCCAATCGTGACATGACGATTAG	*Lactobacillus* phages P1174/Ld3/Ld17/Ld25A
	>spacer49 CTTACTGGTTGGAGCTTTGGCTTCACCGACAGA	*Lactobacillus* phages P1174/phiLdb
	>spacer50 TTGTCGACTGGTACTATATGATAAGACAACGGA	-
	>spacer51 CCAGGCGGAACCGGTCGGCACACCGTTCGTGCA	*Lactobacillus* phage ViSo-2018b
	>spacer52 TTAAAGGGGGAAATCATGAGATATAAGCGAAAA	-
306504..306598	>spacer1 AAATCCGCGGCCAGTTCTGCCGTTAGTTCGACAAAATCC TCCTCAAGCC	*Lb. delbrueckii* subsp. *bulgaricus*
1207296..1207399	>spacer1 CGTCTGCCGCTGGATTATGCCCAAAACGCTTCTTTGCGT AATGGTCATCATG	*Lb. delbrueckii*

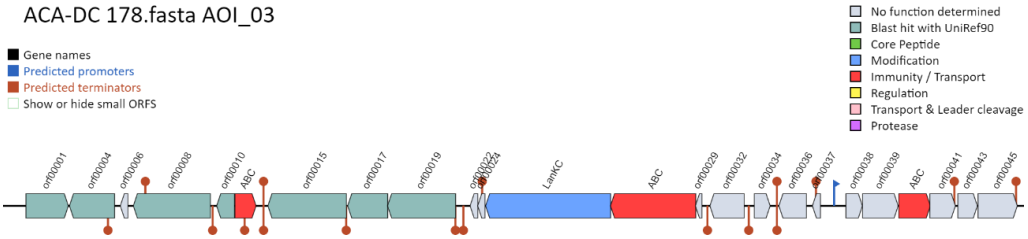


Name	Function
orf00001	Putative acetyltransferase SH0499 OS=Staphylococcus haemolyticus (strain JCSC1435) OX=279808 GN=SH0499 PE=3 SV=1
orf00003	
orf00004	
orf00005	Signal peptidase I OS=Bacillus licheniformis OX=1402 GN=lepB PE=3 SV=1
orf00007	
orf00012	Peptidase T OS=Lactobacillus helveticus OX=1587 GN=pepT PE=1 SV=2
orf00013	Probable transposase for insertion-like sequence element IS1161 OS=Streptococcus salivarius OX=1304 PE=3 SV=1
orf00016	GTP cyclohydrolase 1 type 2 homolog OS=Streptococcus pneumoniae serotype 4 (strain ATCC BAA-334 / TIGR4) OX=170187 GN=SP_1609 PE=1 SV=1
orf00017	GTP cyclohydrolase 1 type 2 homolog OS=Bacillus subtilis (strain 168) OX=224308 GN=yqfO PE=3 SV=2
orf00020	tRNA (adenine(22)-N(1))-methyltransferase OS=Bacillus subtilis (strain 168) OX=224308 GN=trmK PE=1 SV=2
orf00023	Putative transposase in snaA-snaB intergenic region OS=Streptomyces pristinaespiralis OX=38300 PE=3 SV=1
orf00027	Uncharacterized Nudix hydrolase YvcI OS=Bacillus subtilis (strain 168) OX=224308 GN=yvcI PE=3 SV=1
64.3;Enterolysin_A	64.3;Enterolysin_A
orf00031	Methylated-DNA--protein-cysteine methyltransferase OS=Methanosarcina acetivorans (strain ATCC 35395 / DSM 2834 / JCM 12185 / C2A) OX=188937 GN=ogt PE=3 SV=1
orf00033	
orf00035	
orf00036	Putative type I restriction enzyme HindVIIP R protein OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) OX=71421 GN=HI_1285 PE=3 SV=1
orf00039	Putative type I restriction enzyme HindVIIP R protein OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) OX=71421 GN=HI_1285 PE=3 SV=1
orf00040	
orf00042	Type-1 restriction enzyme MjaXIP specificity protein OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) OX=243232 GN=MJ0130 PE=1 SV=3
orf00047	Tyrosine recombinase XerC OS=Pasteurella multocida (strain Pm70) OX=272843 GN=xerC PE=3 SV=1
orf00049	Type-1 restriction enzyme EcoprrI specificity protein OS=Escherichia coli OX=562 GN=prrB PE=3 SV=1
orf00050	Putative type I restriction enzyme HindVIIP M protein OS=Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) OX=71421 GN=HI_1287 PE=3 SV=1
orf00053	Uncharacterized protein MJ0132 OS=Methanocaldococcus jannaschii (strain ATCC 43067 / DSM 2661 / JAL-1 / JCM 10045 / NBRC 100440) OX=243232 GN=MJ0132 PE=3 SV=1

Figure 6.1 Genes encoding for antimicrobial peptides in the genome sequence of *Lb. lactis* ACA-DC 178, as predicted by BAGEL4.



Name	Function	Motifs
orf00001	Probable tRNA-dihydrouridine synthase 1 OS=Bacillus subtilis (strain 168) OX=224308 GN=dus1 PE=3 SV=1	
orf00002	Lysine--tRNA ligase OS=Lactobacillus plantarum (strain ATCC BAA-793 / NCIMB 8826 / WCFS1) OX=220668 GN=lysS PE=3 SV=1	
orf00003		
orf00006	Uncharacterized protein YhcK OS=Bacillus subtilis (strain 168) OX=224308 GN=yhcK PE=4 SV=1	
orf00009	O-acetyl-L-homoserine sulfhydrylase OS=Thermotoga maritima (strain ATCC 43589 / MSB8 / DSM 3109 / JCM 10099) OX=243274 GN=TM_0882 PE=1 SV=1	
orf00010		
orf00012	Protein YtsP OS=Bacillus subtilis (strain 168) OX=224308 GN=ytsP PE=3 SV=2	
orf00013	Chaperone protein ClpB OS=Oceanobacillus iheyensis (strain DSM 14371 / CIP 107618 / JCM 11309 / KCTC 3954 / HTE831) OX=221109 GN=clpB PE=3 SV=1	
42.2;Bovicin_255_peptide	42.2;Bovicin_255_peptide	RBS=TGTATTAGGAGGATATAA
42.2;Bovicin_255_peptide	42.2;Bovicin_255_peptide	RBS=CTTGCAGAATTACAGCAA
EntA_Immun	bacteriocin immunity	PF08951
orf00021	DNA-directed RNA polymerase subunit beta OS=Lactobacillus delbrueckii subsp. bulgaricus (strain ATCC BAA-365) OX=321956 GN=rpoB PE=3 SV=2	
orf00023	DNA-directed RNA polymerase subunit beta' OS=Lactobacillus delbrueckii subsp. bulgaricus (strain ATCC 11842 / DSM 20081 / JCM 1002 / NBRC 13953 / NCIMB 11778) OX=390333 GN=rpoC PE=3 SV=1	
orf00024		
orf00026	5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase YitU OS=Bacillus subtilis (strain 168) OX=224308 GN=yitU PE=1 SV=1	
orf00027	Putative phosphatase YxeH OS=Bacillus subtilis (strain 168) OX=224308 GN=yxeH PE=3 SV=1	
orf00029	30S ribosomal protein S12 OS=Lactobacillus delbrueckii subsp. bulgaricus (strain ATCC 11842 / DSM 20081 / JCM 1002 / NBRC 13953 / NCIMB 11778) OX=390333 GN=rpsL PE=3 SV=1	



Name	Function	Motifs
orf00001	BhtR	
orf00004	ATP-dependent DNA helicase RecG OS=Acidithiobacillus ferrooxidans OX=920 GN=recG PE=3 SV=1	
orf00006		
orf00008	H(+)/Cl(-) exchange transporter ClcA OS=Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696) OX=290338 GN=clcA PE=3 SV=1	
orf00010	Putative transposase InsK for insertion sequence element IS150 OS=Escherichia coli (strain K12) OX=83333 GN=insK PE=2 SV=2	
ABC	Lipoprotein-releasing system ATP-binding protein LolD OS=Dechloromonas aromatica (strain RCB) OX=159087 GN=lolD PE=3 SV=1	PF00005
orf00015	H(+)/Cl(-) exchange transporter ClcA OS=Salmonella gallinarum (strain 287/91 / NCTC 13346) OX=550538 GN=clcA PE=3 SV=1	
orf00017	AbpR [Lactobacillus salivarius subsp. salivarius]	PF00072
orf00019	Bacteriocin production related histidine kinase	
orf00022		
orf00024		
LanKC		lanI;labKC;PF00069
ABC	Multidrug resistance ABC transporter ATP-binding/permease protein BmrA OS=Bacillus subtilis (strain 168) OX=224308 GN=bmrA PE=1 SV=1	PF00005
orf00029		
orf00032		
orf00034		
orf00036		
orf00037		
orf00038		
orf00039		
ABC	Putative lantibiotic ABC transporter,ATP-binding protein precursor	PF00005
orf00041		
orf00043		
orf00045		

Figure 6.1 (continued). Genes encoding for antimicrobial peptides in the genome sequence of *Lb. lactis* ACA-DC 178, as predicted by BAGEL4.

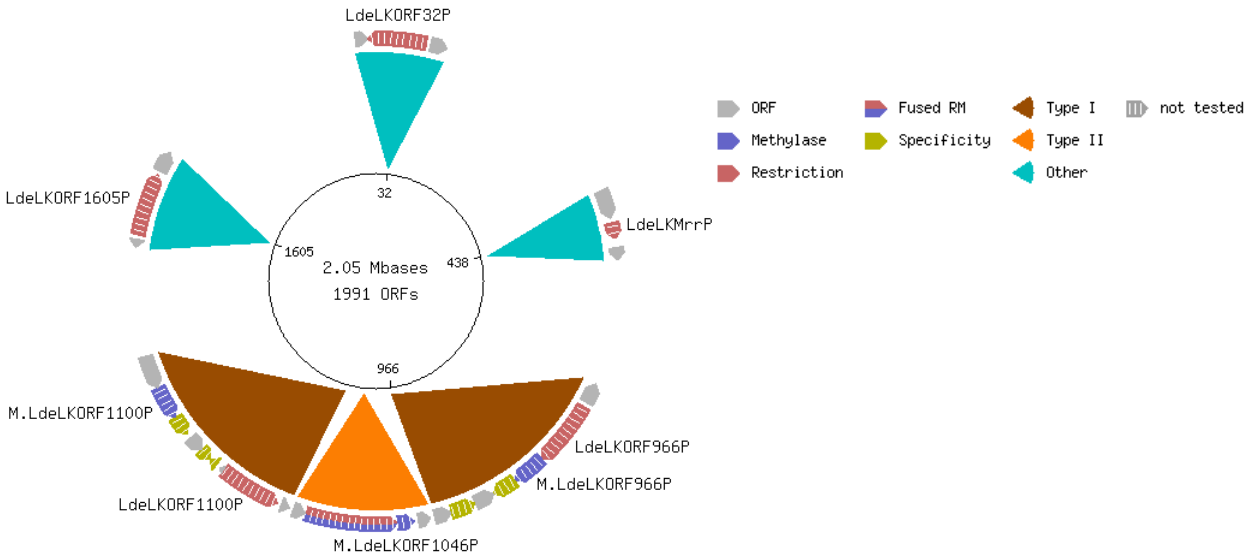
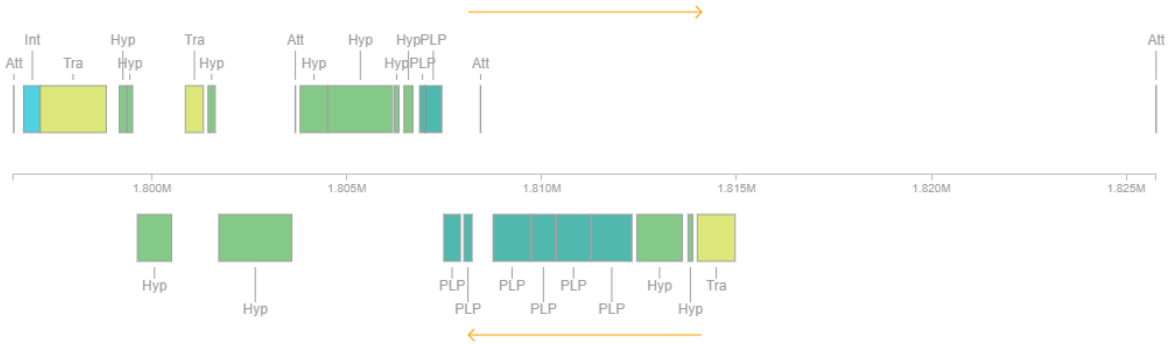


Figure 6.2 Restriction-modification systems predicted in the *Lb. lactis* ACA-DC 178 chromosome by the REBASE database.

Region	Region Length	Completeness	Score	# Total Proteins	Region Position	Most Common Phage	GC %
1	29.3Kb	incomplete	50	23	1796448-1825762	PHAGE_Sphing_PAU_NC_019521(4)	45.68%

Figure 6.3 Incomplete prophage region in the genome sequence of *Lb. lactis* ACA-DC 178, as predicted by PHASTER.



#	CDS Position	BLAST Hit	E-Value
1	1796448..1796460	attL	0.0
2	1796705..1797151	PHAGE_Microc_Ma_LMMO1_NC_008562: putative site-specific integrase-resolvase; PP_01846; phage(gi117530307)	1.62e-46
3	1797120..1798826	PHAGE_Thermu_TMA_NC_015937: transposase; PP_01847; phage(gi343960512)	1.08e-13
4	1799158..1799346	hypothetical; PP_01848	0.0
5	1799361..1799501	hypothetical; PP_01849	0.0
6	complement(1799623..1800501)	hypothetical; PP_01850	0.0
7	1800855..1801319	PHAGE_Ralsto_RSA1_NC_009382: transposase IRSO15-like; PP_01851; phage(gi145708108)	2.53e-06
8	1801436..1801621	hypothetical; PP_01852	0.0
9	complement(1801713..1803587)	hypothetical; PP_01853	0.0
10	1803672..1803683	attL	0.0
11	1803801..1804508	hypothetical; PP_01854	0.0
12	1804519..1806171	hypothetical; PP_01855	0.0
13	1806217..1806336	hypothetical; PP_01856	0.0
14	1806466..1806696	hypothetical; PP_01857	0.0
15	1806865..1807017	PHAGE_Bacill_PBC1_NC_017976: resolvase domain-containing protein; PP_01858; phage(gi389060369)	6.56e-05
16	1807001..1807435	PHAGE_Plankt_PaV_LD_NC_016564: ABC transporter; PP_01859; phage(gi371496158)	4.26e-09
17	complement(1807488..1807913)	PHAGE_Temper_phiNIH1.1_NC_003157: putative methionine sulfoxide reductase; PP_01860; phage(gi16271776)	1.95e-45
18	complement(1808002..1808211)	PHAGE_Rhodoc_E3_NC_021347: putative DNA-binding helix-turn-helix protein; PP_01861; phage(gi509142230)	2.31e-09
19	1808425..1808436	attR	0.0
20	complement(1808752..1809726)	PHAGE_Entero_phi92_NC_023693: Phi92_gp064; PP_01862; phage(gi726646997)	1.06e-36
21	complement(1809740..1810348)	PHAGE_Entero_phi92_NC_023693: Phi92_gp065; PP_01863; phage(gi726646998)	2.62e-55
22	complement(1810375..1811259)	PHAGE_Entero_phi92_NC_023693: Phi92_gp066; PP_01864; phage(gi726646999)	1.11e-119
23	complement(1811279..1812316)	PHAGE_Entero_phi92_NC_023693: Phi92_gp067; PP_01865; phage(gi726647000)	1.97e-104
24	complement(1812439..1813611)	hypothetical; PP_01866	0.0
25	complement(1813755..1813877)	hypothetical; PP_01867	0.0
26	complement(1813995..1814963)	PHAGE_Thermu_TMA_NC_015937: transposase; PP_01868; phage(gi343960512)	1.46e-16
27	1825762..1825774	attR	0.0

Figure 6.4 Incomplete prophage region 1 and BLAST analysis of the respective CDSs in the genome sequence of *Lb. lactis* ACA-DC 178, as predicted by PHASTER.

Table 6.4 EPS gene clusters identified in *Lb. lactis* ACA-DC 178 (based on BlastP similarity searches)

	Locus tag	RefSeq locus tag	Product
EPS cluster 1	ACADC178_1706	ACADC178_RS08920	glycerol-3-phosphate cytidylyltransferase
	ACADC178_1707 (pseudogene)	ACADC178_RS08925	LCP family protein
		ACADC178_RS08930 (pseudogene)	IS30 family transposase
	ACADC178_1710 (pseudogene)	ACADC178_RS08935	PTS transporter subunit EIIC
	ACADC178_1711	ACADC178_RS08940	glycosyltransferase family 2 protein
	ACADC178_1712 (pseudogene)	ACADC178_RS08945	hypothetical protein
	ACADC178_1713	ACADC178_RS08950	transposase
	ACADC178_1714 (pseudogene)	ACADC178_RS08960	hypothetical protein
	ACADC178_1715	ACADC178_RS08965	acyltransferase
	ACADC178_1716	ACADC178_RS08970	Stealth CR1 domain-containing protein
	ACADC178_1717	ACADC178_RS08975	glycosyltransferase
	ACADC178_1718	ACADC178_RS08980	polymerase
	ACADC178_1719	ACADC178_RS08985	glycosyltransferase family 2 protein
	ACADC178_1720	ACADC178_RS08990	DUF4422 domain-containing protein
	ACADC178_1721	ACADC178_RS08995	IS256-like element ISLdl2 family transposase
	ACADC178_1722	ACADC178_RS09000	hypothetical protein
	ACADC178_1723	ACADC178_RS09005	glycosyltransferase family 4 protein
	ACADC178_1724	ACADC178_RS09010	flippase
	ACADC178_1725	ACADC178_RS09015	UDP-galactopyranose mutase
	ACADC178_1726	ACADC178_RS09025	IS30-like element ISL7 family transposase
	ACADC178_1727	ACADC178_RS09030	LCP family protein
EPS cluster 2	ACADC178_1758 (pseudogene)	ACADC178_RS09175	transposase
	ACADC178_1759 (pseudogene)	ACADC178_RS09180	IS607 family transposase
		ACADC178_RS09185 (pseudogene)	transposase
	ACADC178_1762	ACADC178_RS09195	hypothetical protein
	ACADC178_1763	ACADC178_RS09200	flippase
	ACADC178_1764	ACADC178_RS09205	UDP-galactopyranose mutase
	ACADC178_1765	ACADC178_RS09210	glycosyltransferase family 2 protein
	ACADC178_1766 (pseudogene)	ACADC178_RS09215	glycosyltransferase family 4 protein
	ACADC178_1767	ACADC178_RS09220	IS256 family transposase
	ACADC178_1769	ACADC178_RS09225	glycosyltransferase family 2 protein
	ACADC178_1770	ACADC178_RS09230	glycosyltransferase family 4 protein
	ACADC178_1771	ACADC178_RS09235	acyltransferase
	ACADC178_1772	ACADC178_RS09240	glycosyltransferase family 4 protein
	ACADC178_1773	ACADC178_RS09245	DUF1972 domain-containing protein
	ACADC178_1774	ACADC178_RS09250	sugar transferase
	ACADC178_1775	ACADC178_RS09255	exopolysaccharide biosynthesis protein
	ACADC178_1776	ACADC178_RS09260	CpsD/CapB family tyrosine-protein kinase
	ACADC178_1777	ACADC178_RS09265	exopolysaccharide biosynthesis protein
	ACADC178_1778	ACADC178_RS09270	LCP family protein

Table 6.5 Genes implicated in carbohydrate catabolism in *Lb. lactis* ACA-DC 178

Protein	Gene	Locus tag
Galactose metabolism		
Galactokinase	*galK*	ACADC178_1579
Galactose 1-P uridylyltransferase	*galT*	ACADC178_1518
D-glucose transport and metabolism		
PTS system, glucose/glucoside-specific enzyme	-	ACADC178_1627
Glucose-6-phosphate isomerase	*pgi*	ACADC178_0637
Fructose transport and metabolism		
PTS system, fructose-specific enzyme IIABC component	*fruB*	ACADC178_1810
1-phosphofructokinase	*fruk*	ACADC178_1811
Fructose repressor	*fruR*	ACADC178_1812
Fructokinase	*ScrK*	ACADC178_0771
6-Phosphofructokinase	*pfk2*	ACADC178_1592
Fructose-bisphosphate aldolase	*fba*	ACADC178_1403
Mannose transport and metabolism		
PTS system, mannose-specific enzyme IIABCD components	-	ACADC178_1613, ACADC178_1612, ACADC178_1611
Mannose-6-phosphate isomerase	*pmi*	ACADC178_0046
6-Phosphofructokinase	*pfK2*	ACADC178_1592
Fructose-bisphosphate aldolase	*fba*	ACADC178_1403
N-acetylglycosamin transport and metabolism		
N-acetylglucosamine-6-phosphate deacetylase	*nagA*	ACADC178_0197
Glucosamine-6-phosphate deaminase	*nagB*	ACADC178_1930
Salicin transport and metabolism		
6-Phospho-ß-glucosidase	*pbg3*	ACADC178_0283
Glucose-6-phosphate isomerase	-	ACADC178_0637
Maltose transport and metabolism		
Maltose ABC transporter permease protein	*MalE/F/G*	ACADC178_0193, ACADC178_0194, ACADC178_0195
Maltose phosphorylase	*mapA*	ACADC178_0190
ß-phosphoglucomutase	*pgmA*	ACADC178_0191
Maltose ABC transporter permease protein	*malE/F/G*	ACADC178_0193, ACADC178_0194, ACADC178_0195
Oligo-1,6-glucosidase	*malL*	ACADC178_0188
PTS –Glucose specific EIIA	-	ACADC178_1745
Maltose-6'-phosphate glucosidase	*malH*	ACADC178_1631
Protein	**Gene**	**Locus tag**
Lactose transport and metabolism		
PTS family lactose porter, IIABC components	*LacE/LacF*	ACADC178_1575, ACADC178_1577
Transcription antiterminator	*LacT*	ACADC178_1576
6-Phospho-b-galactosidase	*LacG*	ACADC178_1574
Galactose-6-phosphate isomerase LacB subunit	*LacB*	ACADC178_1563

Galactose-6-phosphate isomerase LacA subunit	*LacA*	ACADC178_1564
Tagatose-6-phosphate kinase	*LacC*	ACADC178_1567
Tagatose 1,6-diphosphate aldolase	*LacD*	ACADC178_1568
Lactose permease	*LacS*	ACADC178_1078
ß-galactosidase	*LacZ*	ACADC178_1077
Lac repressor	*LacR*	ACADC178_1076
Saccharose transport and metabolism		
PTS family sucrose porter, EIIBC component	*ScrA*	ACADC178_1591
Sucrose 6P hydrolase (sucrase)	*ScrB*	ACADC178_1590
Fructokinase	*ScrK*	ACADC178_0771
Trehalose transport and metabolism		
Trehalose operon repressor	*treR*	ACADC178_1746
PTS family trehalose porter, IIBC component	*treB*	ACADC178_1745
Trehalose-6-phosphate hydrolase	*treC*	ACADC178_1747
Starch metabolism		
Neopullulanase	*nplT*	ACADC178_0189
Glucan 1,6-alpha -glucosidase	*dexB*	ACADC178_0186

Table 6.6 Genes of the proteolytic system of *Lb. lactis* ACA-DC 178

Proteases and aminopeptidases		
Locus tag	**Gene**	**Protein**
ACADC178_1069	*prtL*	Proteinase precursor
ACADC178_1543	*sms*	Putative ATP-dependent serine protease
ACADC178_1546	*pepC*	Aminopeptidase C
ACADC178_0478	*pepD1*	Dipeptidase
ACADC178_1833	*pepD2*	Dipeptidase
ACADC178_1804	*pepF*	Oligoendopeptidase F
ACADC178_0265	*pepG*	Aminopeptidase G
ACADC178_1671	*pepIP*	Proline iminopeptidase
ACADC178_0519	*pepM*	Methionine aminopeptidase
ACADC178_1860	*pepN*	Aminopeptidase N
ACADC178_0229	*pepO1*	Endopeptidase
ACADC178_1283	*pepP*	X-Pro dipeptidase PepP
ACADC178_1682	*pepPN*	Prolyl aminopeptidase
ACADC178_1428	*pepQ1*	X-Pro dipeptidase PepQ
ACADC178_1085	*pepT1*	Aminotripeptidase (peptidase T)
ACADC178_0346	*pepT2*	Aminotripeptidase (peptidase T)
ACADC178_1560	*pepV*	X-His dipeptidase
ACADC178_0266	*pepW*	Aminopeptidase W
ACADC178_1309 pseudogene	*pepX*	X-Pro dipeptidyl-peptidase
ACADC178_1428	*pepZ*	X-Pro dipeptidase PepZ
ACADC178_1607	*dpp*	Putative dipeptidyl-peptidase
LBUL87_RS02660 pseudogene	*dppA*	D-aminopeptidase
ACADC178_0554		Putative peptidase
ACADC178_0555		Putative peptidase
ACADC178_1458		Putative peptidase
ACADC178_0919		Putative acylaminoacyl-peptidase
ACADC178_1456 pseudogene	*gcp*	Putative glycoprotein endopeptidase
ACADC178_1055	*pcp*	Pyroglutamyl-peptidase I
ACADC178_0355	*clpC*	ATP-dependent Clp protease, ATP-binding subunit
ACADC178_0533	*clpE*	ATP-dependent Clp protease, ATP-binding subunit ClpE
ACADC178_1877	*clpL*	ATP-dependent Clp protease, ATP-binding subunit
ACADC178_0581	*clpP*	ATP-dependent Clp protease proteolytic subunit
ACADC178_0724	*clpX*	ATP-dependent Clp protease ATP-binding subunit clpX
ACADC178_1149	*hslU, clpY*	ATP-dependent protease HslVU (ClpYQ), ATPase subunit
ACADC178_1150	*hslV, clpQ*	ATP-dependent protease HslVU (ClpYQ), peptidase subunit
ACADC178_0157	*htpX1*	Protease
ACADC178_1063	*htpX2*	Protease
ACADC178_1411	*dacA*	D-alanyl-D-alanine carboxypeptidase
ACADC178_0139	*htrA*	HtrA-like serine protease

ACADC178_1394	*PrsA or PrtM*	protease maturation protein
ACADC178_1083		signal peptidase I
ACADC178_0928		signal peptidase II

Peptide and amino acid transporters

Locus tag	Gene	Protein
ACADC178_0254	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_0255	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_0259	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_0260	*oppB*	Oligopeptide ABC transporter, permease protein
ACADC178_0261	*oppC*	Oligopeptide ABC transporter, permease protein
ACADC178_0262	*oppD*	Oligopeptide ABC transporter, ATP-binding protein
ACADC178_0263	*oppF*	Oligopeptide ABC transporter, ATP-binding protein
ACADC178_1247	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_1248	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_1249	*oppC*	Oligopeptide ABC transporter, permease protein
ACADC178_1250	*oppB*	Oligopeptide ABC transporter, permease protein
ACADC178_1251	*oppF*	Oligopeptide ABC transporter, ATP-binding protein
ACADC178_1252	*oppD*	Oligopeptide ABC transporter, ATP-binding protein
ACADC178_0345	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_0409	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_0730	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_1002	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_1520	*oppA*	Oligopeptide ABC transporter, substrate binding protein
ACADC178_0484	*glnQ*	Glutamine ABC transporter, ATP-binding protein
ACADC178_0485	*glnH1*	Glutamine ABC transporter, glutamine-binding protein
ACADC178_0486	*glnM*	Glutamine ABC transporter, permease protein
ACADC178_0487	*glnP*	Glutamine ABC transporter, permease protein
ACADC178_0488	*glnH2*	Glutamine ABC transporter, glutamine-binding protein
ACADC178_1945	-	Glutamine ABC transporter, substrate binding / permease protein
ACADC178_1946 pseudogene	-	Glutamine ABC transporter, ATP-binding protein
ACADC178_1059	-	Glycine betaine ABC transporter, substrate binding / permease protein
ACADC178_1060	-	Glycine betaine ABC transporter, ATP-binding protein
ACADC178_0164	-	Amino Acid ABC transporter, ATP-binding protein
ACADC178_0165 pseudogene	-	Amino acid ABC transporter, substrate binding / permease protein
ACADC178_0892	-	Amino Acid ABC transporter, permease protein
ACADC178_0893	-	Amino acid ABC transporter, permease protein
ACADC178_0894	-	Amino acid ABC transporter, ATP-binding protein
ACADC178_0895	-	Amino acid ABC transporter, substrate binding protein
ACADC178_1173	-	Amino acid ABC transporter, ATP-binding protein
ACADC178_1174	-	Amino Acid ABC transporter, substrate binding /permease protein
ACADC178_0293	-	Amino acid permease, APC family

ACADC178_0505	-	Amino acid permease, APC family
ACADC178_0514	-	Amino acid permease, APC family
ACADC178_1325	-	Amino acid permease, APC family
ACADC178_1535	-	Amino acid permease, APC family
ACADC178_1561	-	Amino acid permease, APC family
ACADC178_1608	-	Amino acid permease
ACADC178_0457 pseudogene	*brnQ*	Branched-chain amino acid transport system carrier protein
ACADC178_0485	-	Glutamine-binding protein
ACADC178_0286	-	Amino acid ABC transporter, substrate binding protein
ACADC178_0287	-	Amino acid ABC transporter, ATP-binding protein
ACADC178_0288	-	Amino acid ABC transporter, permease protein
ACADC178_0599	-	Spermidine/putrescine ABC transporter, permease protein
ACADC178_0600	-	Spermidine/putrescine ABC transporter, permease protein
ACADC178_0601 pseudogene	-	Spermidine/putrescine ABC transporter, substrate binding protein
ACADC178_0736	-	Methionine ABC transporter, substrate binding protein
ACADC178_0737	-	Methionine ABC transporter, ATP-binding protein
ACADC178_0738	-	Methionine ABC transporter, permease protein
ACADC178_1689	-	Amino acid ABC transporter, substrate binding protein
ACADC178_1690	-	Amino acid ABC transporter, ATP-binding protein
ACADC178_1691	-	Amino acid ABC transporter, permease protein
ACADC178_0821	-	Threonine/serine exporter family protein
ACADC178_1797	-	Threonine/serine exporter family protein
ACADC178_1798	-	Threonine/serine exporter family protein

7

Concluding remarks and future perspectives

Lactic acid bacteria are the kings among starters used in fermented foods, and at the same time comprise the majority of probiotic microorganisms. In particular, *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Lactobacillus delbrueckii* subsp. *lactis* are among the most significant starters used worldwide in dairy fermented foods, such as fermented milks, yogurt and diverse cheese varieties. Thus, understanding their functionality during milk fermentation is of major significance.

Over the last years, whole-genome sequencing emerged as a comprehensive method for analyzing entire genomes, and, in combination with microbial bioinformatics analysis and comparative genomics, rendered the prediction of function of most genes in a bacterial genome feasible. In addition, valuable insights about the bacterial evolution have been gained from the identification of newly acquired or lost genetic elements within a genome. This type of genomic analysis allowed the identification of strain-specific metabolic properties related to acidification ability and production of various technologically important molecules (EPS, antimicrobial metabolites, flavor and aroma compounds, health beneficial molecules). These findings enabled the correlation between genetic variation and phenotypic diversity at the strain level to a degree, which may facilitate the selection of suitable strains towards the optimization of microbial metabolic performance during the fermentation process and, thus, the overall quality of the final product. Moreover, the amelioration of sequencing technologies, the continuous refinement of bioinformatics tools for sequence data analysis, the proliferation of biological databases and repositories and the wide availability of and accessibility to biological information rendered genomic analysis to an efficacious tool for the in-depth study of food related bacteria. Nonetheless, the total assignment of genotypic traits to technological properties that would allow the selection of a starter culture directly from *in sililco* data is yet to be established and further improvement in sequencing and analysis tools is mandatory.

In this context, the aim of the present thesis was the evaluation at the genome level of three LAB strains of the ACA-DC culture collection of the Laboratory of Dairy Research of the Agricultural University of Athens, namely *S. thermophilus* ACA-DC 2, *Lb. delbrueckii subsp. bulgaricus* ACA-DC 87 and *Lb. delbrueckii subsp. lactis* ACA-DC 178, isolated from naturally fermented Greek dairy products. For this purpose, the complete chromosomal sequences of the

three strains were obtained using next generation sequencing techniques and sequence assembly was evaluated through whole-genome optical mapping and whole-genome alignments. Advanced bioinformatics methods were employed for structural and functional annotation as well as for comparative genomics analysis either among strains within each species or among strains of closely related species.

The genome sequence of *S. thermophilus* ACA-DC 2 has been shaped through reductive evolution, as attested by its small genome size. Milk niche-related adaptive traits include the presence of a large number of putative pseudogenes, several of which are implicated in carbohydrate metabolism and virulence. Genomic analysis of *S. thermophilus* ACA-DC 2 revealed the existence of several genomic features desirable for dairy foods production. These include one complete lactose-galactose operon and one exopolysaccharide cluster. The strain also carries several proteolytic enzymes, although it lacks the cell-envelope proteinase PrtS. Its absence is probably indicative of low acidifying ability but does not compromise its potential for protocooperation with *Lb. bulgaricus*. Stress response genes have also been identified, which suggests that the strain can cope with adverse conditions during food production. Although the strain has one putative CRISPR-cas system with only one spacer, indicating low activity or inactivation of the cas proteins, the presence of putative R-M systems may contribute to resistance against phages invasion. In addition, the identification of putative antimicrobial peptides may enhance further the competitiveness of the strain in food ecosystems and the potential elimination of dairy pathogens.

Comparative genomic analysis among 23 *S. thermophilus strains* with complete genomes available in the NCBI database showed a high degree of conservation at the genome level. However, strain-depended differences related to gene content, presence of certain genomic features, genome size and architecture were also identified. Detailed pan genome analysis revealed that the core genome of the species comprises several genes implicated in essential metabolic processes, such as lactose and urea metabolism, peptide degradation, amino acid and peptide transport, amino acid biosynthesis, while a number of unique genes were also identified in several strains. Analysis also showed that the species pan genome is still open, thus, the sequencing of additional strains, especially those found in niches other than milk and dairy

products, will facilitate the comprehension of the species evolution. However, it seems that the species gene pool is rather limited, thus, its pan genome could be soon characterized as closed. This observation is rather rationale if taken into account that the lifestyle of the species, which is narrowed to dairy habitats, limits its access to the global microbial gene pool of the various ecological ecosystems.

Core genome and ANI phylogenetic analyses revealed that the clustering of *S. thermophilus* strains was abide by a discrete pattern based on certain genomic features. Therefore, cluster A comprised strains with larger genomes, the majority of which encode *prtS* and carry the His biosynthetic gene cluster along with certain CRISPR-Cas system types and specific GIs, in contrast to the strains included in cluster B. These findings pointed out the existence of at least two major lineages within the species. Thorough investigation showed the existence of subgroups within these clusters, which was also supported on a variable degree by COG analysis and the presence/absence of specific loci and/or their organization, such as EPS clusters, CRISPR arrays, R-M systems and GIs. Finally, loci essential for the physiology and the metabolism of the species, such as lactose/galactose and urease operon, aminopeptidases, protocooperation related genes, the majority of amino acid and peptide transporters, as well as most of the amino acid biosynthetic pathways, were found to be conserved in all strains. From a technological point of view, several strain-dependent traits related to proteolysis, amino acid transport and metabolism, EPS production and defense mechanisms, were also identified. These findings clearly reinforce our knowledge towards the implementation of suitable *S. thermophilus* strains with desirable traits in food production.

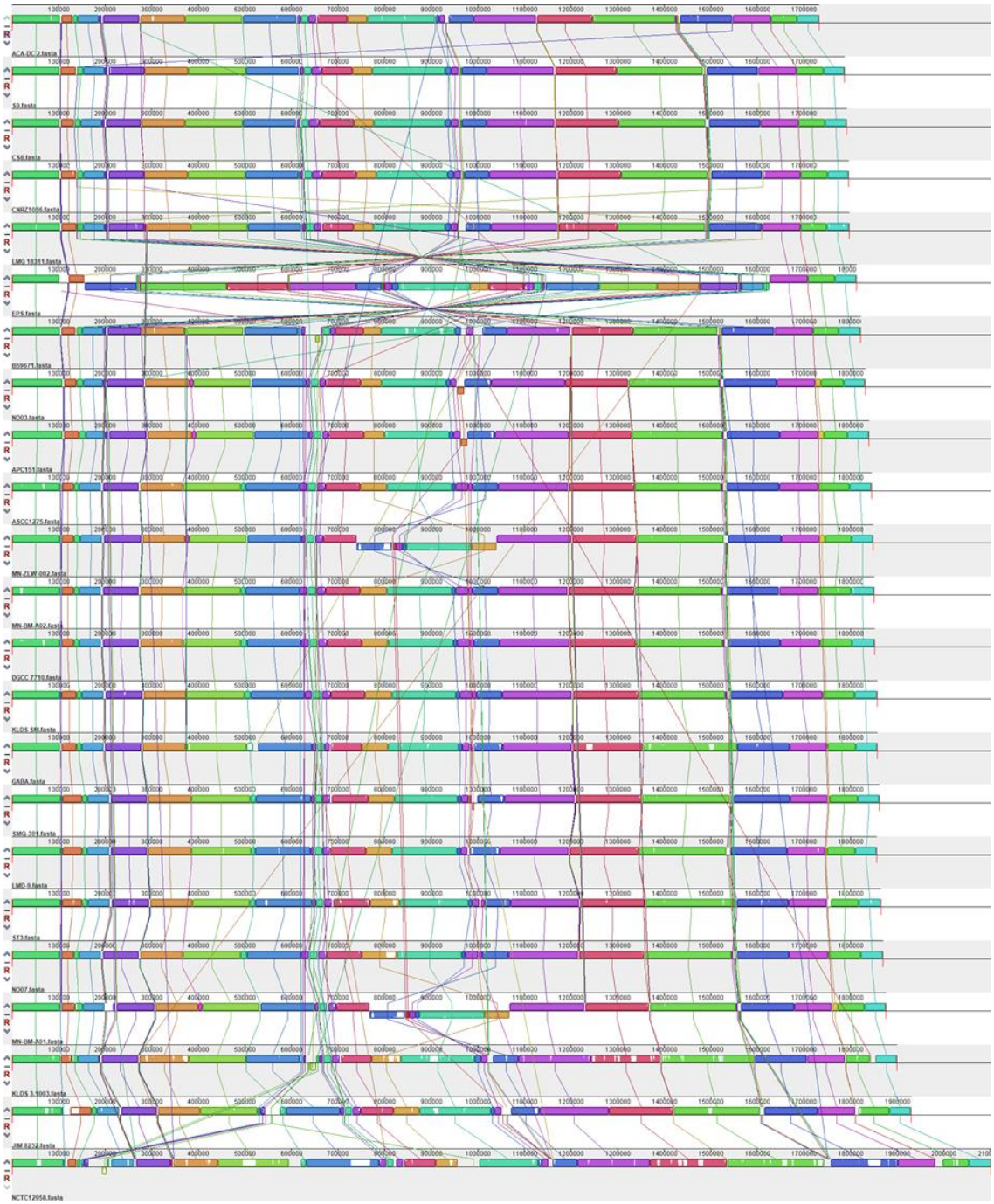
Genomic analysis applied for the genomes of *Lb. delbrueckii* subsp*. bulgaricus* ACA-DC 87 *and Lb. delbrueckii* subsp*. lactis* ACA-DC 178, including evaluation of genome assembly, as well as structural and functional genome annotation indicated the presence of basic technological-related genomic properties. Both subspecies carry one CEP implicated in casein degradation, two distinct EPS gene clusters for EPS production, and defense mechanisms machinery, including one CRISPR-Cas, several R-M systems and antimicrobial peptides. Comparative genomics analysis not only among strains within each subspecies but also among strains of the species *Lb. delbrueckii*

as a whole will provide important information concerning the technological potential of these strains, pinpointing their suitability as starter or adjunct cultures in food fermentations.

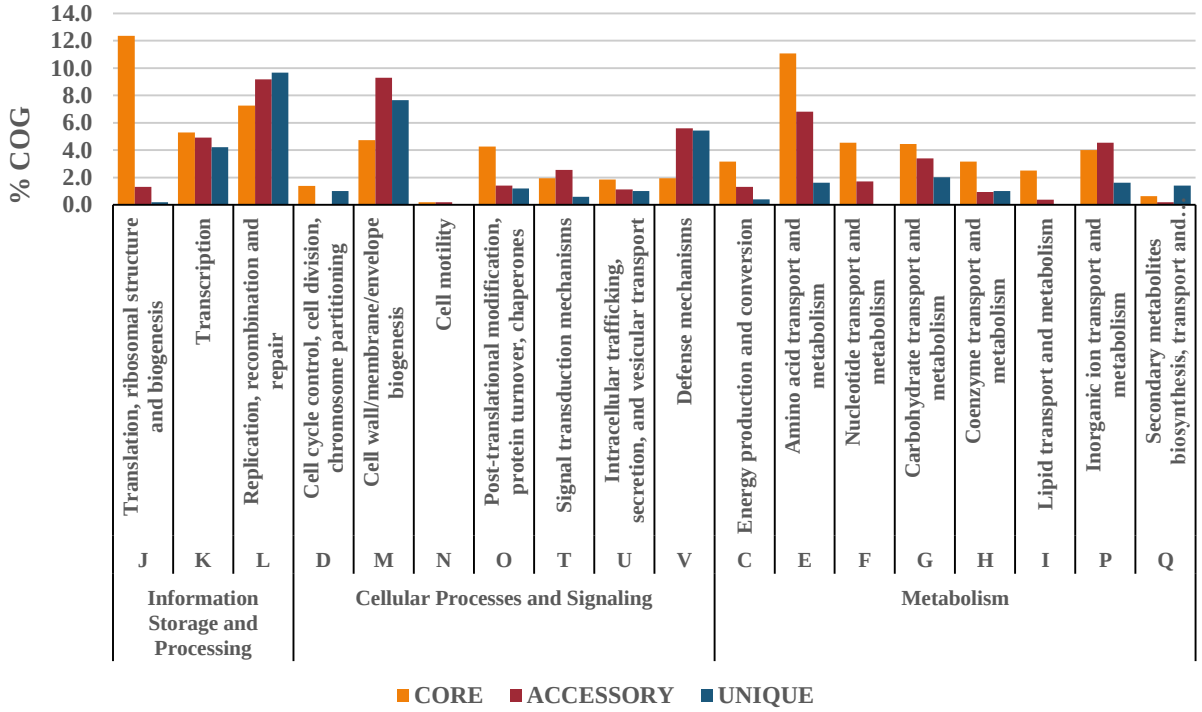
The continuous accumulation of complete genome sequences of these important dairy bacteria enhances the discriminating and informative value of whole-genome and comparative genomics analyses, towards the improved understanding of the physiology, metabolism and evolution of each strain/species. Such comprehensive research provides genetic-based knowledge about desirable phenotypes, application-depended suitability, adaptability, virulence and probiotic efficacy of the strains under investigation.

Appendix

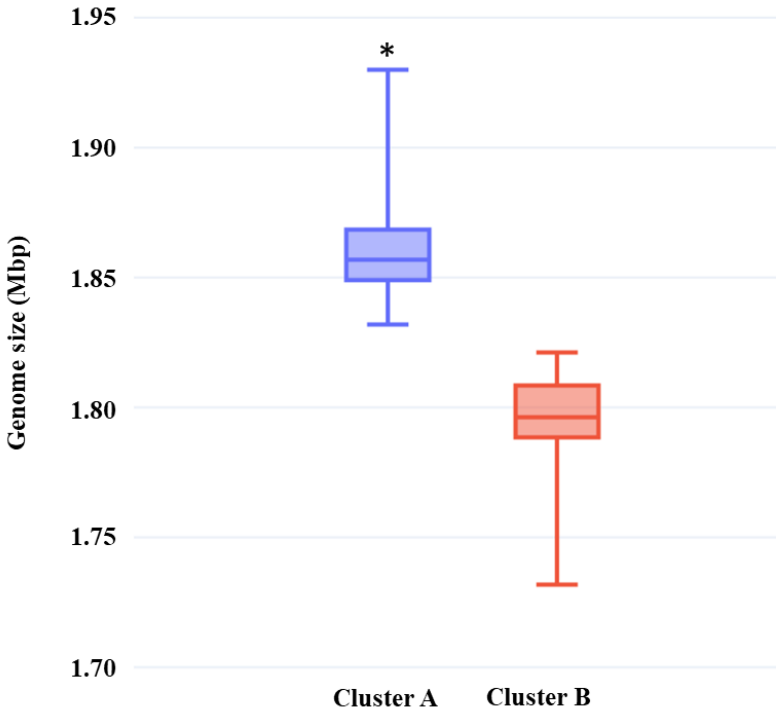
Supplementary Material - Chapter 3 - article Alexandraki V., Kazou M., Blom J., Pot B., Papadimitriou K., Tsakalidou E., (2019) Comparative genomics of *Streptococcus thermophilus* support important traits concerning the evolution, biology and technological properties of the species. Front Microbiol, 10:2916. doi: 10.3389/fmicb.2019.02916



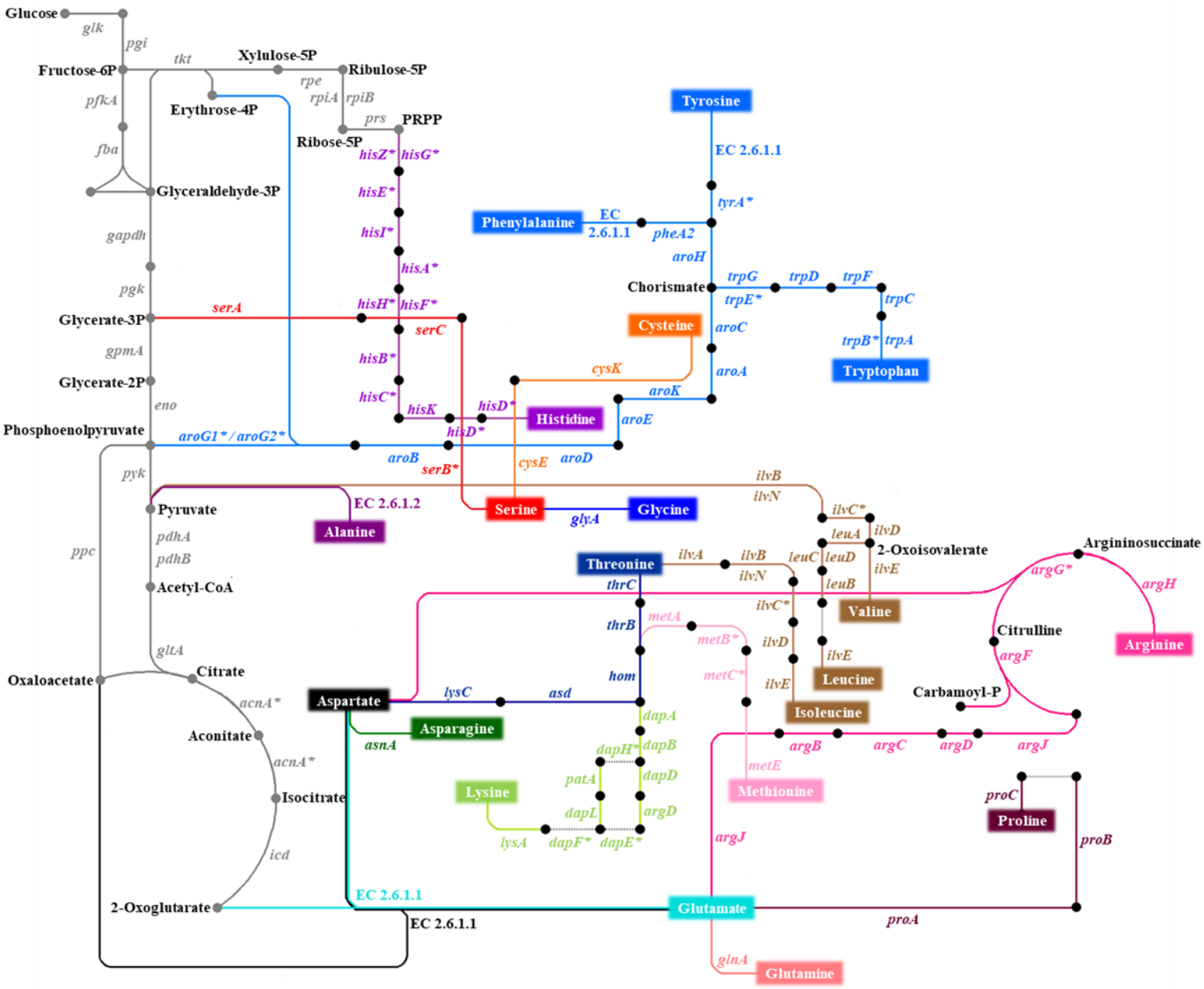
Supplementary Figure 1. Chromosome alignments of the 23 *S. thermophilus* strains with complete genomes as calculated by progressive Mauve. Local collinear blocks (LCBs) of conserved sequences among the strains are represented by rectangles of the same color. Connecting lines can be used to visualize synteny or rearrangement. LCBs positioned above or under the chromosome (blackline) correspond to the forward and reverse orientation, respectively. The level of conservation is equivalent to the level of vertical color filling within the LCBs (e.g. white regions are strain-specific). Sequences not placed within an LCB are unique for the particular strain. All genomes were synchronized from the *dnaA*



Supplementary Figure 2. Distribution of COG categories (%) for the core, accessory and unique genes identified in the 23 *S. thermophilus* strains

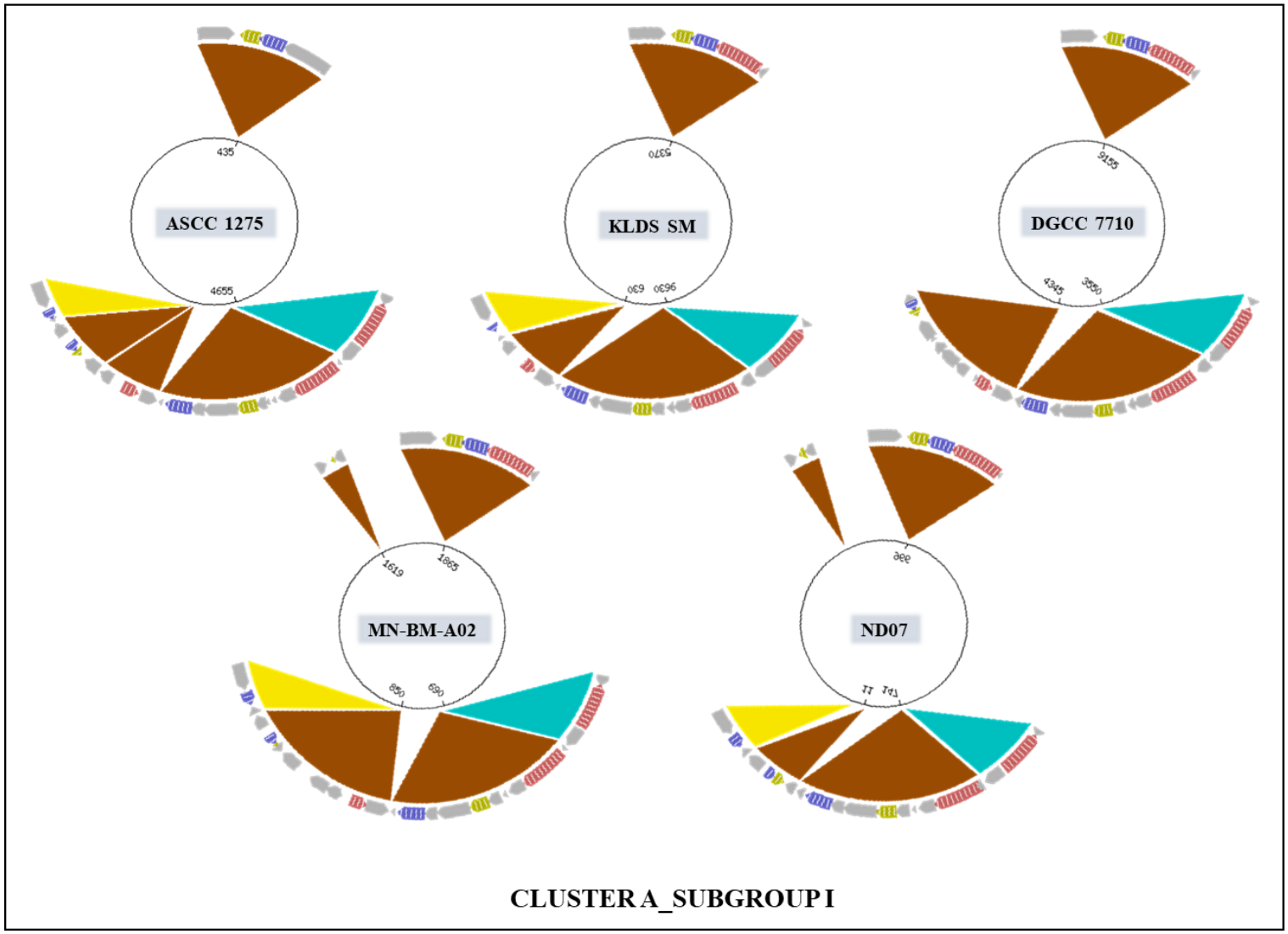


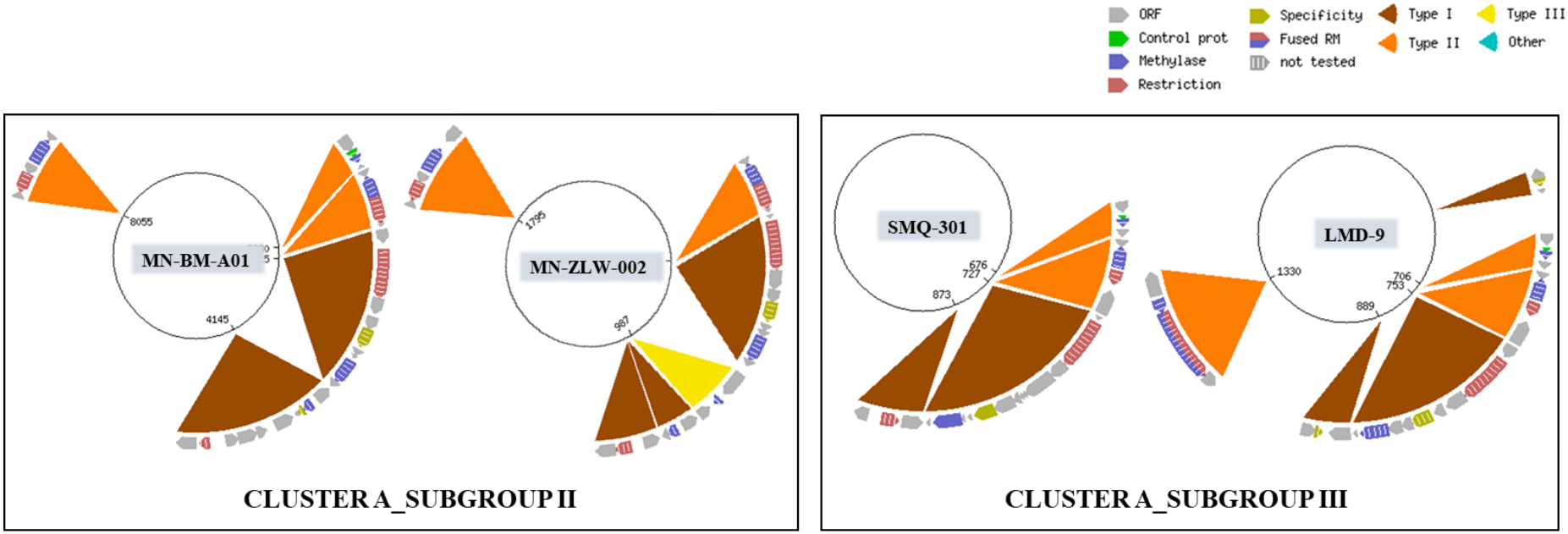
Supplementary Figure 3. Genome size distributions of cluster A and cluster B *S. thermophilus* strains analyzed in this study. The asterisk indicates significant difference between the two clusters as supported by the Mann-Whitney U Test ($p<0.05$).


Glucose
glk
pgi
Fructose-6P
tkt
Xylulose-5P
Ribulose-5P
rpe
rpiA
rpiB
pfkA
Erythrose-4P
prs
PRPP
Ribose-5P
fba
hisZ*
hisG*
hisE*
Glyceraldehyde-3P
hisI*
gapdh
hisA*
pgk
hisH*
hisF*
serA
Glycerate-3P
serC
gpmA
hisB*
Glycerate-2P
hisC*
eno
hisK
hisD*
Histidine
Phosphoenolpyruvate
hisD*
aroG1*/aroG2*
aroB
serB*
cysE
aroD
pyk
Pyruvate
EC 2.6.1.2
Alanine
Serine
glyA
Glycine
ppc
pdhA
pdhB
Acetyl-CoA
gltA
Oxaloacetate
Citrate
acnA*
Aconitate
acnA*
Isocitrate
icd
2-Oxoglutarate
EC 2.6.1.1
EC 2.6.1.1
Glutamate
proA
glnA
Glutamine
Tyrosine
EC 2.6.1.1
tyrA*
Phenylalanine
EC 2.6.1.1
pheA2
aroH
Chorismate
trpG
trpD
trpF
trpE*
trpC
Cysteine
aroC
trpB*
trpA
cysK
aroA
Tryptophan
aroK
aroE
ilvB
ilvN
ilvC*
leuA
ilvD
2-Oxoisovalerate
ilvE
Threonine
ilvA
ilvB
ilvN
leuC
leuD
leuB
thrC
Valine
ilvC*
thrB
metA
metB*
ilvD
ilvE
hom
metC*
ilvE
Leucine
Isoleucine
Aspartate
lysC
asd
dapA
Asparagine
asnA
dapH*
dapB
patA
dapD
Lysine
dapL
argD
lysA
dapF*
dapE*
metE
Methionine
argJ
argB
argC
argD
argJ
Carbamoyl-P
Argininosuccinate
argG*
argH
Citrulline
argF
Arginine
proC
Proline
proB

Supplementary Figure 4. Biosynthesis of amino acids in the 23 *S. thermophilus* strains. Genes represented with solid lines are present in *S. thermophilus* genomes analyzed in this study while genes represented in dashed lines are absent from all strains. Lines highlighted in light grey correspond to spontaneous reactions. Genes with an asterisk are either possible pseudogenes or absent for at least one of the strains analyzed. Our analysis suggests complete biosynthetic pathways in all strains for aspartate, threonine, glycine, proline, glutamine, asparagine, alanine, phenylalanine, cysteine and glutamate. Incomplete biosynthetic pathways were detected for: lysine in all strains, serine and tyrosine in strain B59671, methionine in strains B59671, NCTC12958^{T} and ST3, histidine in strains ACA-DC 2, B59671, CNRZ1066, CS8, EPS, LMG 18311 and S9, arginine in strain MN-BM-A01, branched-chain amino acids in strain JIM 8232 and tryptophan in strain EPS. Pathways for amino acids biosynthesis were compiled from Kyoto Encyclopedia of Genes and Genomes (KEGG) database and Hols et al. (2005)

Supplementary Figure 5. Restriction-modification (R-M) system in the 23 *S. thermophilus* strains as predicted by REBASE database

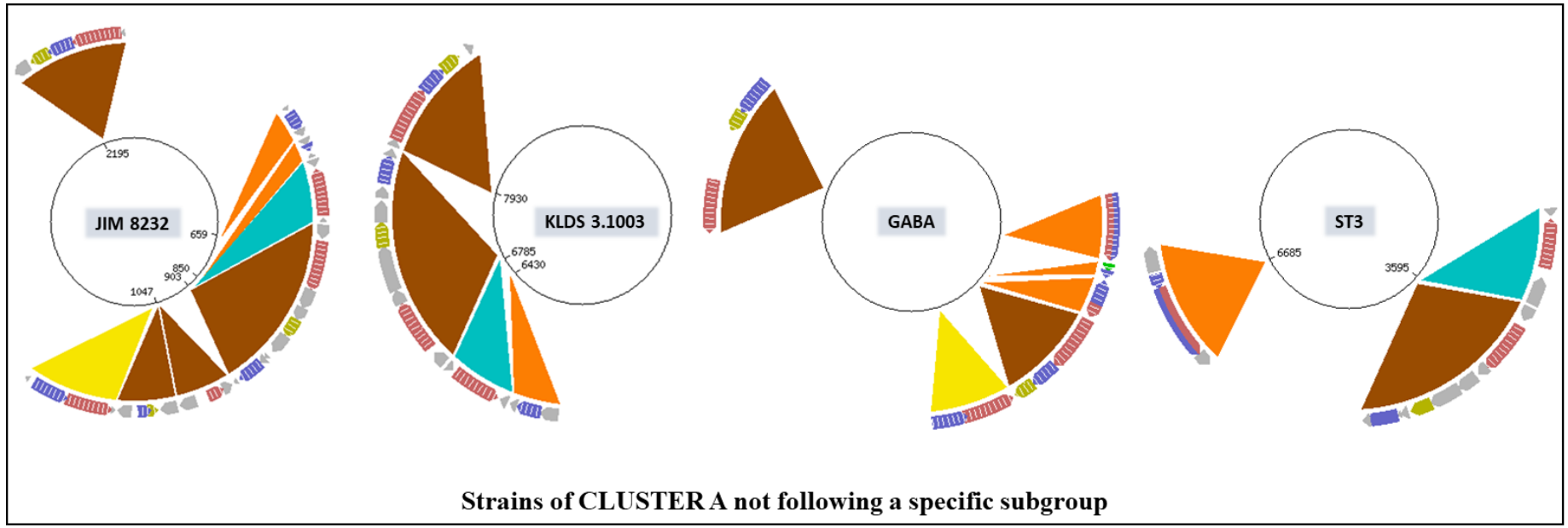




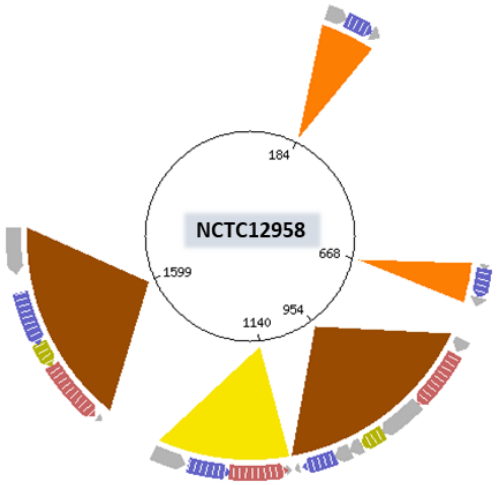
CLUSTER A_SUBGROUP II

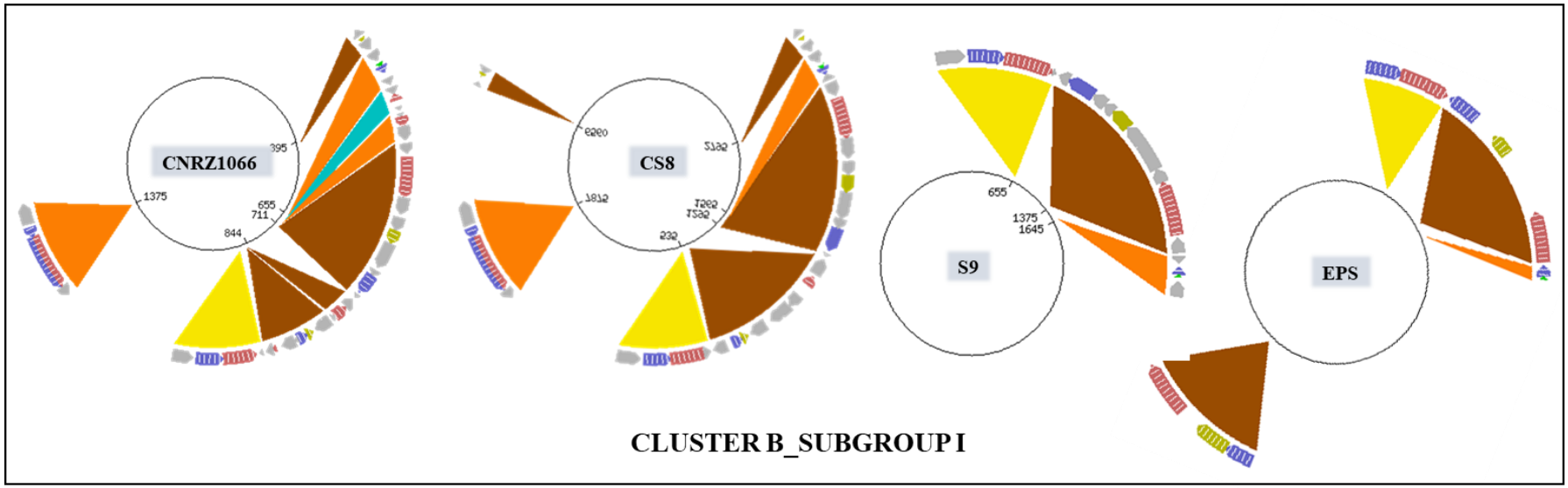
CLUSTER A_SUBGROUP III

CLUSTER A_SUBGROUP IV

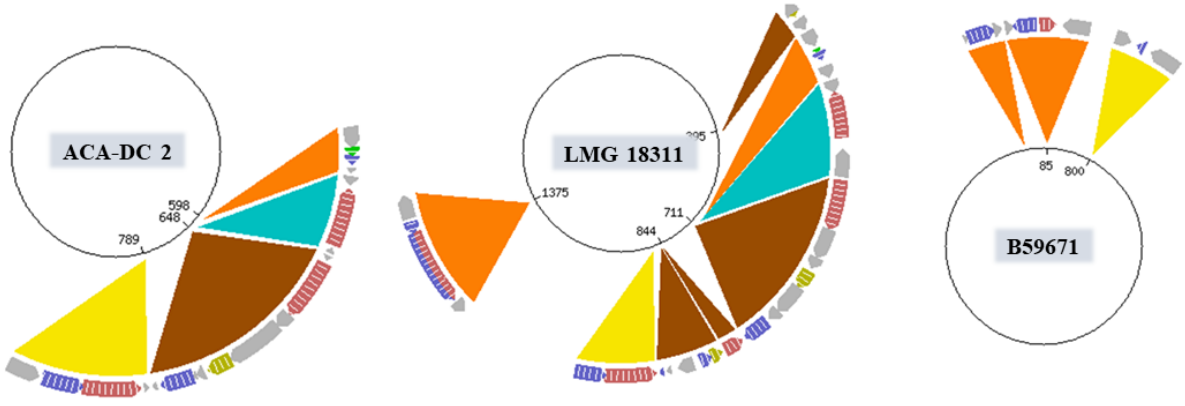


Strains of CLUSTER A not following a specific subgroup



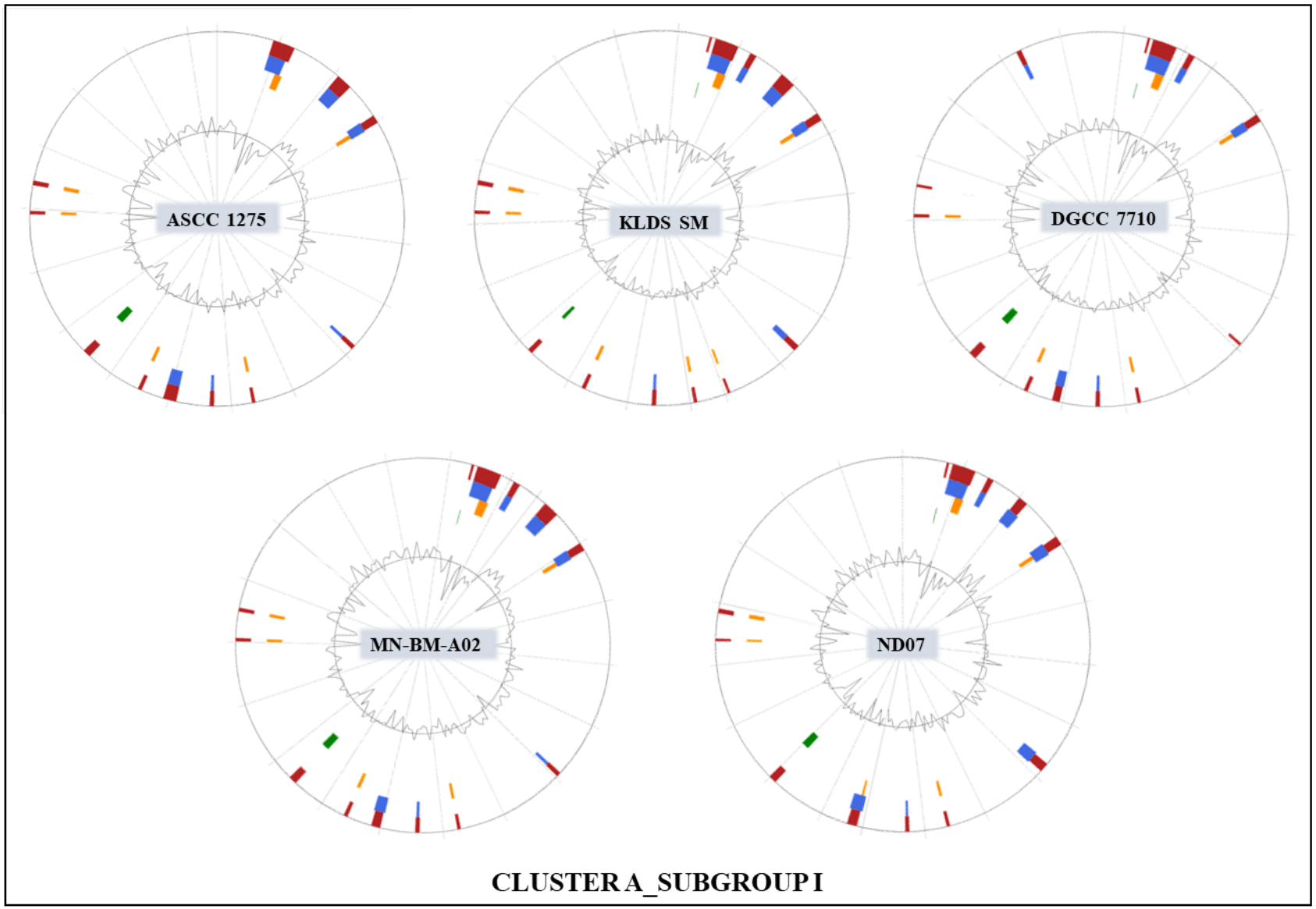


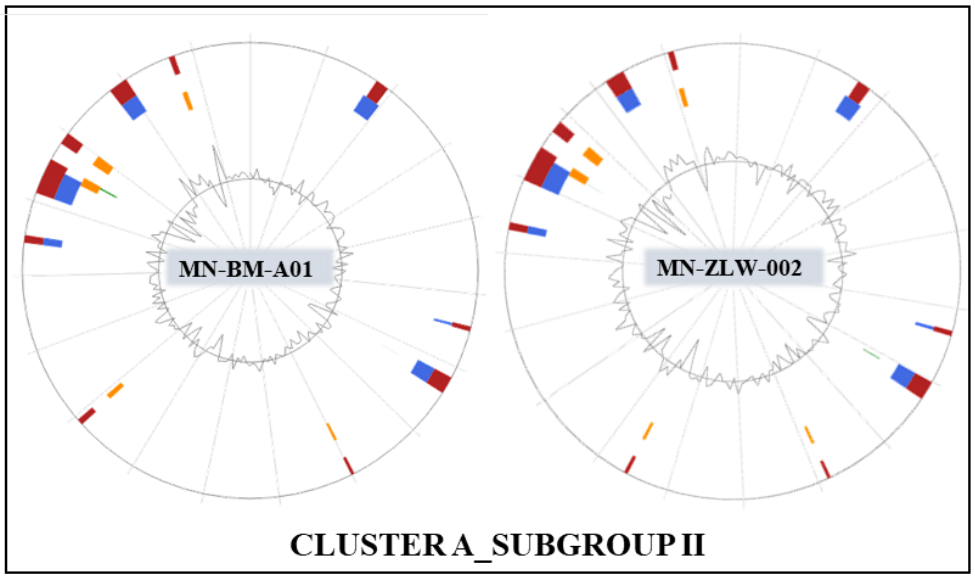
CLUSTER B_SUBGROUP I



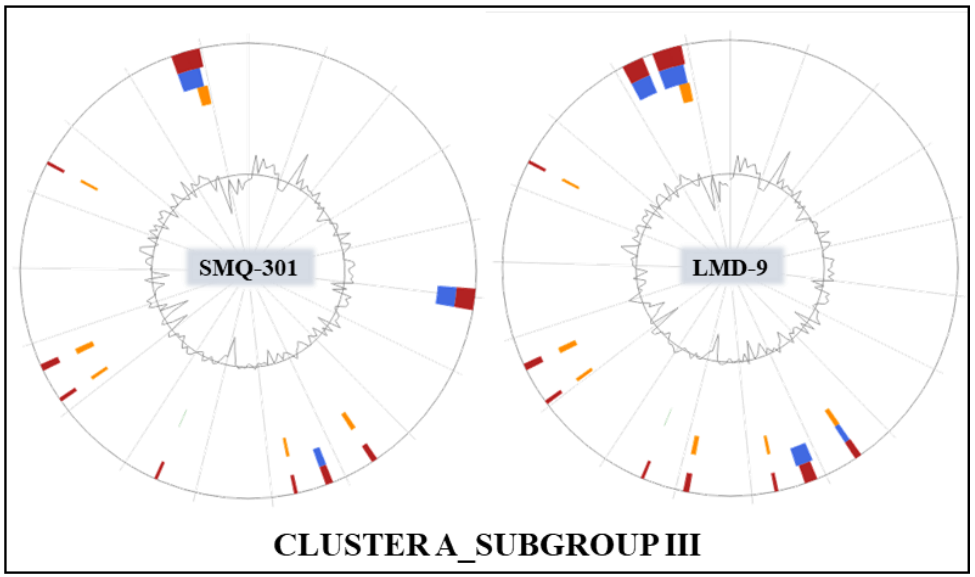
Strains of CLUSTER B not following a specific subgroup

Supplementary Figure 6. Genomic islands (Gis) in the 23 *S. thermophilus* strains as predicted by IslandViewer 4

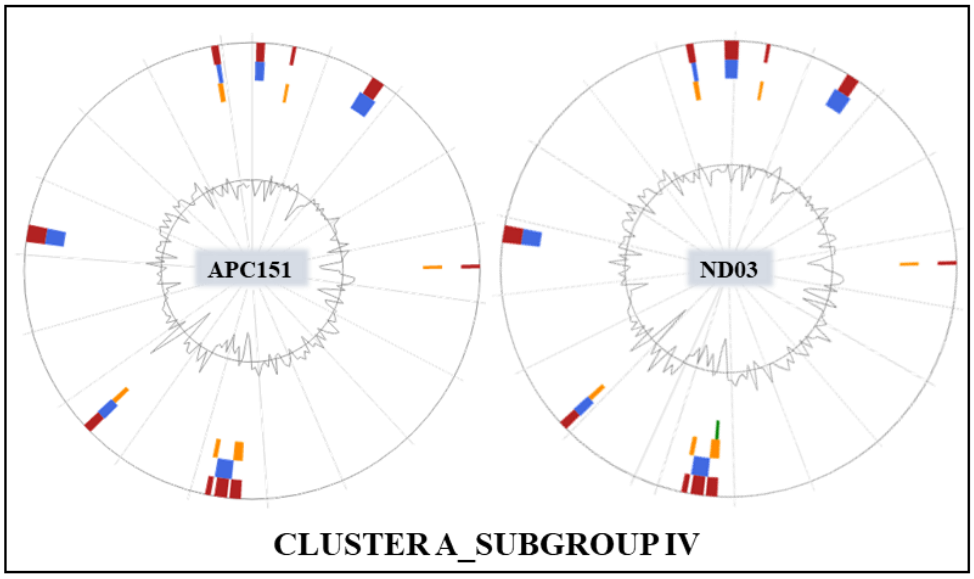




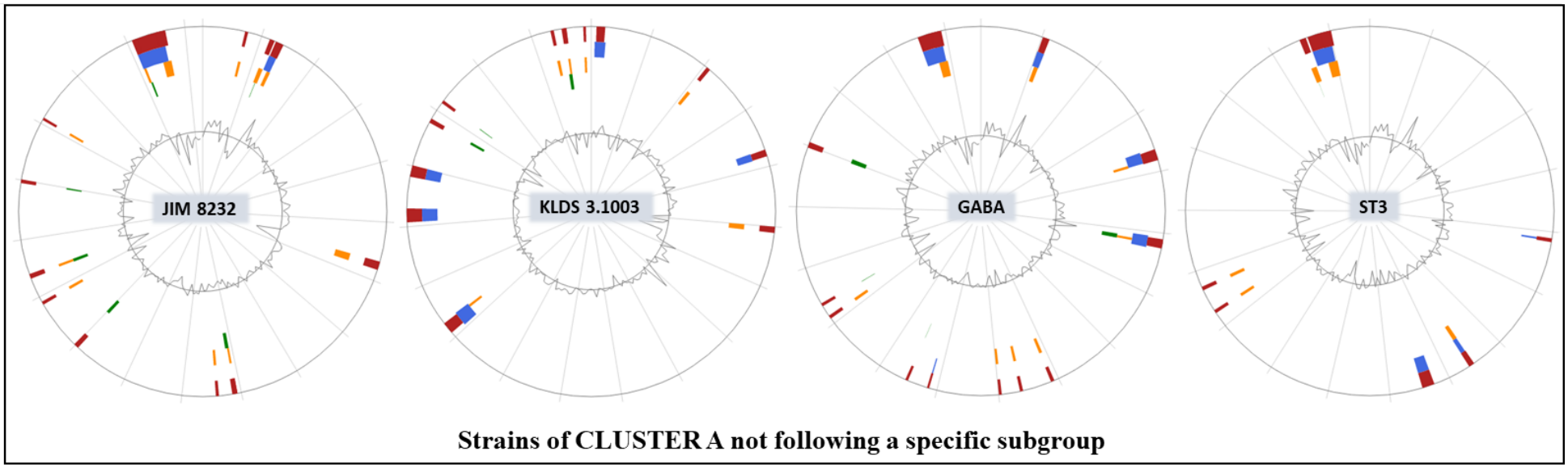
CLUSTER A_SUBGROUP II



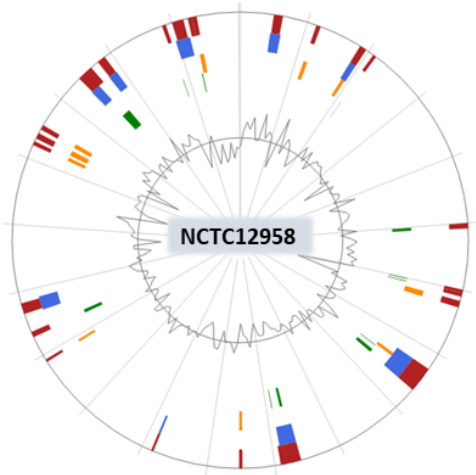
CLUSTER A_SUBGROUP III

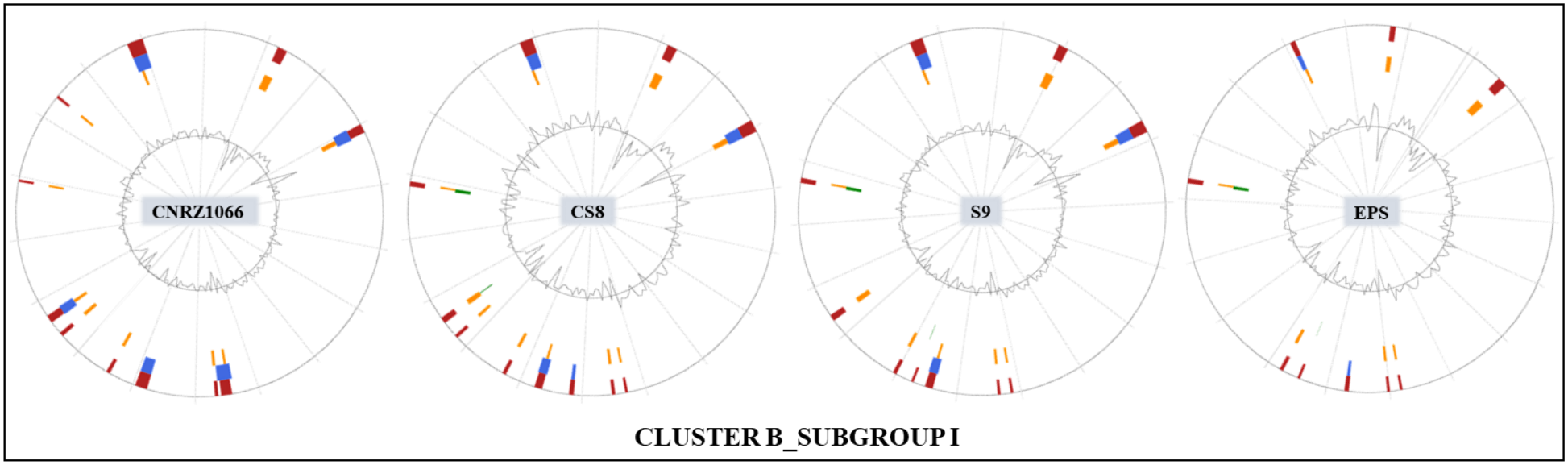


CLUSTER A_SUBGROUP IV

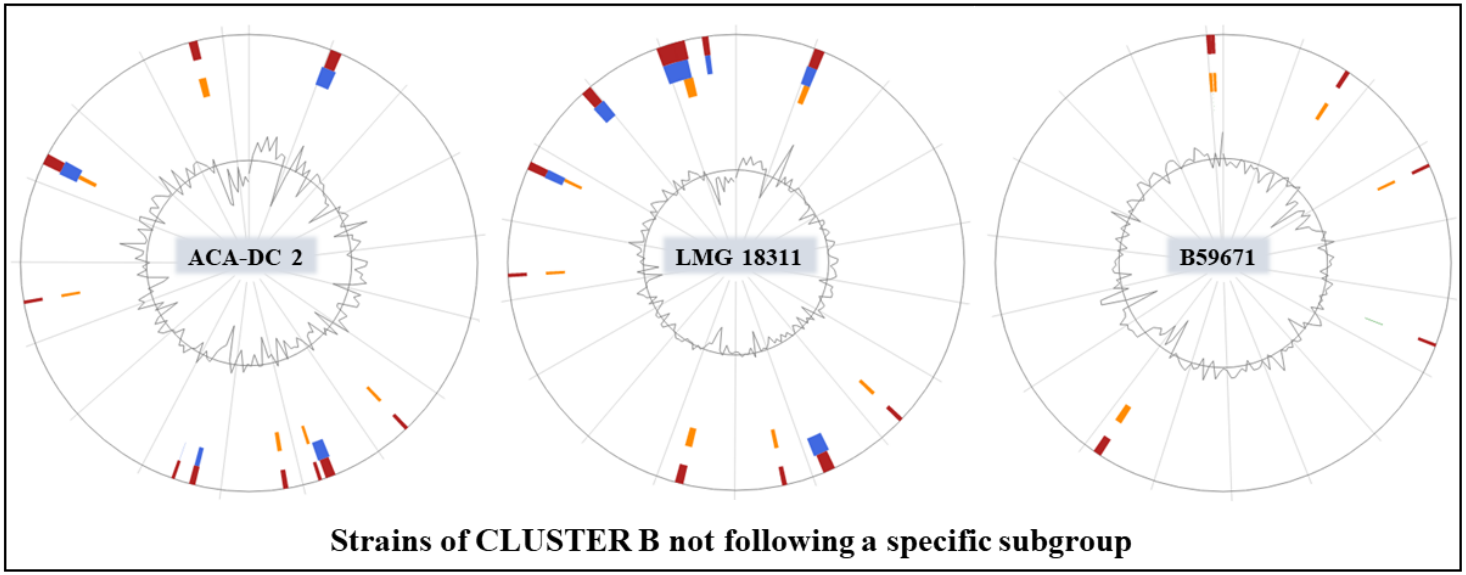


Strains of CLUSTER A not following a specific subgroup





CLUSTER B_SUBGROUP I



Strains of CLUSTER B not following a specific subgroup

Supplementary Table S1. Number of accessory, unique and exclusively absent genes resulted from pan genome analysis of the 23 *S. thermophilus* strains presented according to the clusters and subgroups described in the text

Cluster	Subgroup	Strain	Genes in accessory genome	Unique genes	Exclusively absent genes
No cluster		NCTC12958^T	501	204	31
A	No subgroup	KLDS 3.1003	519	41	13
	I	ASCC 1275	532	0	4
		ND07	541	0	2
		DGCC 7710	526	3	0
		KLDS SM	538	1	1
		MN-BM-A02	539	0	1
	II	MN-ZLW-002	564	1	0
		MN-BM-A01	522	0	28
	No subgroup	JIM 8232	556	67	6
	III	LMD-9	535	7	5
		SMQ-301	536	13	0
	IV	ND03	568	0	0
		APC151	562	0	4
	No subgroup	GABA	485	17	3
	No subgroup	ST3	501	12	7
B	**I**	CNRZ1066	525	0	3
		CS8	526	0	2
		S9	513	0	0
		EPS	481	15	13
	No subgroup	LMG 18311	483	20	7
	No subgroup	B59671	433	22	29
	No subgroup	ACA-DC 2	432	14	10

Supplementary Table S2. Unique genes identified in the 23 *S. thermophilus* strains

Strain (Number of unique genes)	locus_tag	product
KLDS SM (1)	A9497_RS10550	hypothetical protein
MN-ZLW-002 (1)	Y1U_RS04305	glycosyltransferase family 92 protein
DGCC 7710 (3)	CW339_RS06830	hypothetical protein
	CW339_RS06825	hypothetical protein
	CW339_RS03100	hypothetical protein
LMD-9 (7)	STER_RS05265	polysaccharide polymerase
	STER_RS09860	restriction endonuclease subunit S
	STER_RS05280	glycosyltransferase family 1 protein
	STER_RS05270	glycosyl transferase family 1
	STER_RS05260	cell wall biosynthesis glycosyltransferase
	STER_RS05275	cell wall biosynthesis glycosyltransferase
	STER_RS05285	multidrug MFS transporter
ST3 (12)	BGL51_RS03670	ATP-binding protein
	BGL51_RS03665	hypothetical protein
	BGL51_RS02905	glutathione S-transferase
	BGL51_RS02900	glutathione-dependent reductase
	BGL51_RS02890	amino acid ABC transporter substrate-binding protein
	BGL51_RS03660	hypothetical protein
	BGL51_RS04370	HXXEE domain-containing protein
	BGL51_RS02895	N-acetyltransferase
	BGL51_RS09265	CHAP domain-containing protein
	BGL51_RS09270	DNA-binding protein
	BGL51_RS06725	hypothetical protein
	BGL51_RS04340	hypothetical protein
SMQ-301 (13)	SMQ301_RS03660	ATP-binding protein
	SMQ301_RS03675	ATP-binding protein
	SMQ301_RS05280	hypothetical protein
	SMQ301_RS03680	restriction endonuclease subunit S
	SMQ301_RS05285	hypothetical protein
	SMQ301_RS03655	hypothetical protein
	SMQ301_RS05290	hypothetical protein

	SMQ301_RS05300	sugar transferase
	SMQ301_RS09995	hypothetical protein
	SMQ301_RS10850	hypothetical protein
	SMQ301_RS10430	hypothetical protein
	SMQ301_RS03665	hypothetical protein
	SMQ301_RS01095	peptide ABC transporter permease
ACA-DC 2 (14)	STACADC2_RS05025	hypothetical protein
	STACADC2_RS03520	hypothetical protein
	STACADC2_RS06850	methylated-DNA--[protein]-cysteine S-methyltransferase
	STACADC2_RS04100	restriction endonuclease subunit S
	STACADC2_RS05010	hypothetical protein
	STACADC2_RS05015	hypothetical protein
	STACADC2_RS06740	hypothetical protein
	STACADC2_RS05045	sugar transferase
	STACADC2_RS00970	peptide ABC transporter ATP-binding protein
	STACADC2_RS05005	hypothetical protein
	STACADC2_RS04635	hypothetical protein
	STACADC2_RS05020	hypothetical protein
	STACADC2_RS02990	oxidoreductase
	STACADC2_RS04860	hypothetical protein
EPS (15)	CR922_RS00645	DEAD/DEAH box helicase
	CR922_RS00635	restriction endonuclease subunit S
	CR922_RS00630	type I restriction-modification system subunit M
	CR922_RS00650	ATP-dependent DNA helicase
	CR922_RS00640	anticodon nuclease
	CR922_RS00595	site-specific integrase
	CR922_RS00610	hypothetical protein
	CR922_RS05485	XRE family transcriptional regulator
	CR922_RS00600	XRE family transcriptional regulator
	CR922_RS00620	transposase
	CR922_RS00625	transposase
	CR922_RS04445	hypothetical protein
	CR922_RS00615	hypothetical protein
	CR922_RS00605	DNA-binding protein
	CR922_RS04785	hypothetical protein

GABA (17)	CR921_RS02755	hypothetical protein
	CR921_RS02745	type II restriction endonuclease
	CR921_RS07850	restriction endonuclease
	CR921_RS07855	SIR2 family protein
	CR921_RS07865	N-6 DNA methylase
	CR921_RS04420	type III restriction endonuclease subunit M
	CR921_RS07200	hypothetical protein
	CR921_RS07195	transcriptional regulator
	CR921_RS02800	ltrA group II intron reverse transcriptase/maturase
	CR921_RS07860	restriction endonuclease
	CR921_RS03665	restriction endonuclease subunit S
	CR921_RS02115	hypothetical protein
	CR921_RS07190	glycosyl transferase
	CR921_RS02765	hypothetical protein
	CR921_RS02110	cytoplasmic protein
	CR921_RS02120	hypothetical protein
	CR921_RS04740	hypothetical protein
LMG 18311 (20)	STU_RS14565	exopolysaccharide biosynthesis protein
	STU_RS14560	exopolysaccharide biosynthesis protein
	STU_RS14550	glycosyltransferase family 1 protein
	STU_RS14570	alpha-1,2-fucosyltransferase
	STU_RS14605	polysaccharide pyruvyl transferase family protein
	STU_RS14545	exopolysaccharide biosynthesis protein, glycosyl transferase
	STU_RS14575	exopolysaccharide biosynthesis protein
	STU_RS14590	exopolysaccharide biosynthesis protein, glycosyltransferase
	STU_RS14555	hypothetical protein
	STU_RS18955	acyltransferase
	STU_RS14585	hypothetical protein
	STU_RS14615	multidrug MFS transporter
	STU_RS14610	serine acetyltransferase
	STU_RS14595	hypothetical protein
	STU_RS11805	hypothetical protein
	STU_RS19675	hypothetical protein
	STU_RS18960	hypothetical protein
	STU_RS18965	glycosyltransferase family 2 protein

	STU_RS13670	transposase
	STU_RS19480	hypothetical protein
B59671 (22)	CG712_RS00975	LysM peptidoglycan-binding domain-containing protein
	CG712_RS08285	CHAP domain-containing protein
	CG712_RS01340	cmoA carboxy-S-adenosyl-L-methionine synthase CmoA
	CG712_RS01755	hypothetical protein
	CG712_RS01335	MFS transporter
	CG712_RS01720	hypothetical protein
	CG712_RS01760	glycosyltransferase family 2 protein
	CG712_RS01730	hypothetical protein
	CG712_RS01770	capsule biosynthesis protein CapG
	CG712_RS01765	glycosyl transferase family 2
	CG712_RS09515	XRE family transcriptional regulator
	CG712_RS09500	hypothetical protein
	CG712_RS09495	hypothetical protein
	CG712_RS09465	hypothetical protein
	CG712_RS01725	hypothetical protein
	CG712_RS06320	hypothetical protein
	CG712_RS05005	NADPH:quinone reductase
	CG712_RS00160	hypothetical protein
	CG712_RS01690	hypothetical protein
	CG712_RS09525	ComC/BlpC family peptide pheromone/bacteriocin
	CG712_RS09335	hypothetical protein
	CG712_RS09510	hypothetical protein
KLDS 3.1003 (41)	BEN15_RS10575	hypothetical protein
	BEN15_RS02770	hypothetical protein
	BEN15_RS02785	ABC transporter permease
	BEN15_RS06655	recombinase family protein
	BEN15_RS02765	hypothetical protein
	BEN15_RS00420	hypothetical protein
	BEN15_RS00460	hypothetical protein
	BEN15_RS07915	amino acid permease
	BEN15_RS07940	restriction endonuclease subunit S
	BEN15_RS00950	MFS transporter
	BEN15_RS02760	hypothetical protein

	BEN15_RS07920	histidine decarboxylase, pyruvoyl type
	BEN15_RS01005	FRG domain-containing protein
	BEN15_RS08225	sugar ABC transporter permease
	BEN15_RS08545	IS4/IS5 family transposase
	BEN15_RS06515	ABC transporter permease
	BEN15_RS01015	hypothetical protein
	BEN15_RS02780	ABC transporter ATP-binding protein
	BEN15_RS06510	peptide ABC transporter ATP-binding protein
	BEN15_RS08290	cell filamentation protein Fic
	BEN15_RS02750	TetR/AcrR family transcriptional regulator
	BEN15_RS07855	XRE family transcriptional regulator
	BEN15_RS02755	hypothetical protein
	BEN15_RS02800	hypothetical protein
	BEN15_RS02775	HXXEE domain-containing protein
	BEN15_RS07910	hypothetical protein
	BEN15_RS05190	XRE family transcriptional regulator
	BEN15_RS02790	hypothetical protein
	BEN15_RS01010	hypothetical protein
	BEN15_RS06585	bacteriocin
	BEN15_RS09925	hypothetical protein
	BEN15_RS02795	XRE family transcriptional regulator
	BEN15_RS06595	hypothetical protein
	BEN15_RS05195	XRE family transcriptional regulator
	BEN15_RS07875	hypothetical protein
	BEN15_RS06610	ComC/BlpC family peptide pheromone/bacteriocin
	BEN15_RS06590	bacteriocin BlpN
	BEN15_RS05170	hypothetical protein
	BEN15_RS06615	hypothetical protein
	BEN15_RS01325	hypothetical protein
	BEN15_RS06500	YydF family exported signaling peptide
JIM 8232 (67)	STH8232_RS00690	KR domain-containing protein
	STH8232_RS00810	hypothetical protein
	STH8232_RS04765	DUF927 domain-containing protein
	STH8232_RS00680	beta-ketoacyl synthase
	STH8232_RS03020	ATP-dependent endonuclease

	STH8232_RS03025	ATP-dependent helicase
	STH8232_RS00540	alkyl hydroperoxide reductase subunit F
	STH8232_RS09385	hypothetical protein
	STH8232_RS03060	hypothetical protein
	STH8232_RS00760	MFS transporter
	STH8232_RS04775	site-specific integrase
	STH8232_RS04470	MFS transporter
	STH8232_RS00685	aminotransferase class V-fold PLP-dependent enzyme
	STH8232_RS00485	replication initiation protein
	STH8232_RS03065	MFS transporter
	STH8232_RS05550	polysaccharide polymerase
	STH8232_RS00470	site-specific integrase
	STH8232_RS09390	hypothetical protein
	STH8232_RS03055	pyridoxal-dependent decarboxylase
	STH8232_RS07735	radical SAM/SPASM domain-containing protein
	STH8232_RS00745	cell division protein FtsK
	STH8232_RS07730	radical SAM protein
	STH8232_RS04760	hypothetical protein
	STH8232_RS09415	toll/interleukin-1 receptor domain-containing protein
	STH8232_RS00765	NAD(P)-dependent alcohol dehydrogenase
	STH8232_RS00700	ABC transporter ATP-binding protein
	STH8232_RS07725	helix-turn-helix domain-containing protein
	STH8232_RS00705	ABC transporter permease
	STH8232_RS00795	XRE family transcriptional regulator
	STH8232_RS00805	hypothetical protein
	STH8232_RS02595	hypothetical protein
	STH8232_RS00695	hypothetical protein
	STH8232_RS10600	hypothetical protein
	STH8232_RS00755	hypothetical protein
	STH8232_RS04745	zinc ribbon domain-containing protein
	STH8232_RS00775	HAD family hydrolase
	STH8232_RS00785	histidine phosphatase family protein
	STH8232_RS00500	glutamine amidotransferase
	STH8232_RS00710	TetR/AcrR family transcriptional regulator
	STH8232_RS00740	hypothetical protein

	STH8232_RS10075	restriction endonuclease subunit S
	STH8232_RS00735	plasmid replication protein
	STH8232_RS04735	recombinase
	STH8232_RS00720	flavodoxin family protein
	STH8232_RS00545	peroxiredoxin
	STH8232_RS03050	flavoprotein
	STH8232_RS05520	transposase
	STH8232_RS00790	hypothetical protein
	STH8232_RS00505	DUF1697 domain-containing protein
	STH8232_RS04755	DUF1643 domain-containing protein
	STH8232_RS00490	hypothetical protein
	STH8232_RS00495	XRE family transcriptional regulator
	STH8232_RS00770	Rrf2 family transcriptional regulator
	STH8232_RS00480	hypothetical protein
	STH8232_RS00780	carboxymuconolactone decarboxylase family protein
	STH8232_RS05490	acyltransferase
	STH8232_RS00750	hypothetical protein
	STH8232_RS04730	hypothetical protein
	STH8232_RS04740	hypothetical protein
	STH8232_RS09410	hypothetical protein
	STH8232_RS05365	hypothetical protein
	STH8232_RS00675	peptidase S51
	STH8232_RS02950	hypothetical protein
	STH8232_RS00660	DNA-binding protein
	STH8232_RS09400	hypothetical protein
	STH8232_RS00475	DUF3173 domain-containing protein
	STH8232_RS04770	DUF3173 domain-containing protein
NCTC12958 (204)	DQL34_RS04100	phage tail protein
	DQL34_RS05315	LPXTG cell wall anchor domain-containing protein
	DQL34_RS04110	hypothetical protein
	DQL34_RS09750	DUF3991 domain-containing protein
	DQL34_RS05340	type IV secretory system conjugative DNA transfer family protein
	DQL34_RS07660	HsdR family type I site-specific deoxyribonuclease
	DQL34_RS01070	virulence associated protein
	DQL34_RS02700	CHAP domain-containing protein

	DQL34_RS07670	type I restriction-modification system subunit M
	DQL34_RS05270	hypothetical protein
	DQL34_RS02695	DUF87 domain-containing protein
	DQL34_RS04555	hypothetical protein
	DQL34_RS05350	hypothetical protein
	DQL34_RS03290	ATP-binding cassette domain-containing protein
	DQL34_RS04115	hypothetical protein
	DQL34_RS05370	hypothetical protein
	DQL34_RS05295	AAA family ATPase
	DQL34_RS09765	type IV secretory system conjugative DNA transfer family protein
	DQL34_RS05290	type IA DNA topoisomerase
	DQL34_RS09755	DNA topoisomerase III
	DQL34_RS09735	relaxase
	DQL34_RS01020	ATP-binding cassette domain-containing protein
	DQL34_RS01170	recombinase
	DQL34_RS01165	recombinase
	DQL34_RS04105	phage tail protein
	DQL34_RS04040	phage portal protein
	DQL34_RS05200	replication protein RepR
	DQL34_RS01175	recombinase
	DQL34_RS03275	sensor histidine kinase
	DQL34_RS09715	ISLre2 family transposase
	DQL34_RS05380	CHAP domain-containing protein
	DQL34_RS05410	ISLre2 family transposase
	DQL34_RS01025	hypothetical protein
	DQL34_RS07665	restriction endonuclease subunit S
	DQL34_RS03075	PTS beta-glucoside transporter subunit IIBCA
	DQL34_RS04035	PBSX family phage terminase large subunit
	DQL34_RS10465	site-specific integrase
	DQL34_RS03890	site-specific integrase
	DQL34_RS05325	beta-hexosaminidase
	DQL34_RS10505	IS3 family transposase
	DQL34_RS09210	isocitrate/isopropylmalate dehydrogenase family protein
	DQL34_RS05305	hypothetical protein
	DQL34_RS04060	major capsid protein E

	DQL34_RS09185	ArsR family transcriptional regulator
	DQL34_RS05235	hypothetical protein
	DQL34_RS03960	DUF1351 domain-containing protein
	DQL34_RS00560	LD-carboxypeptidase
	DQL34_RS09780	hypothetical protein
	DQL34_RS09220	4-hydroxy-2-oxovalerate aldolase
	DQL34_RS09190	EamA family transporter
	DQL34_RS03295	ATP-binding cassette domain-containing protein
	DQL34_RS05230	hypothetical protein
	DQL34_RS03915	DUF3102 domain-containing protein
	DQL34_RS01075	hypothetical protein
	DQL34_RS02660	replication initiator protein A
	DQL34_RS02720	LPXTG cell wall anchor domain-containing protein
	DQL34_RS04045	hypothetical protein
	DQL34_RS02785	cell filamentation protein Fic
	DQL34_RS00725	ABC transporter ATP-binding protein
	DQL34_RS03940	DnaD domain protein
	DQL34_RS04135	peptidoglycan hydrolase
	DQL34_RS05240	hypothetical protein
	DQL34_RS09200	hypothetical protein
	DQL34_RS05390	cobyric acid synthase CobQ
	DQL34_RS01065	DUF1837 domain-containing protein
	DQL34_RS02775	ParA family protein
	DQL34_RS09725	peptide transporter
	DQL34_RS05260	DNA-entry nuclease
	DQL34_RS10785	SDR family NAD(P)-dependent oxidoreductase
	DQL34_RS03945	DNA replication protein
	DQL34_RS05310	LPXTG cell wall anchor domain-containing protein
	DQL34_RS09195	DUF4931 domain-containing protein
	DQL34_RS09215	alpha/beta hydrolase
	DQL34_RS00370	DUF3800 domain-containing protein
	DQL34_RS03300	transporter
	DQL34_RS03305	ABC transporter permease
	DQL34_RS09815	hypothetical protein
	DQL34_RS03280	DNA-binding response regulator

	DQL34_RS04020	DUF1340 domain-containing protein
	DQL34_RS05190	DUF2786 domain-containing protein
	DQL34_RS09790	hypothetical protein
	DQL34_RS09205	DlpA protein
	DQL34_RS03955	single-stranded DNA-binding protein
	DQL34_RS00730	ABC-2 transporter permease
	DQL34_RS05365	hypothetical protein
	DQL34_RS04050	DUF4355 domain-containing protein
	DQL34_RS03245	transcriptional regulator
	DQL34_RS02005	cysteine hydrolase
	DQL34_RS02770	recombinase family protein
	DQL34_RS05385	hypothetical protein
	DQL34_RS08735	DUF1211 domain-containing protein
	DQL34_RS03205	CPBP family intramembrane metalloprotease
	DQL34_RS03985	DUF1642 domain-containing protein
	DQL34_RS00700	XRE family transcriptional regulator
	DQL34_RS07715	hypothetical protein
	DQL34_RS02765	hypothetical protein
	DQL34_RS04145	elongation factor G
	DQL34_RS03920	hypothetical protein
	DQL34_RS02725	hypothetical protein
	DQL34_RS00720	hypothetical protein
	DQL34_RS04085	phage major tail protein, TP901-1 family
	DQL34_RS03990	DUF1642 domain-containing protein
	DQL34_RS10495	lspA signal peptidase II
	DQL34_RS05180	modification methylase Sau96I
	DQL34_RS04030	terminase small subunit
	DQL34_RS05345	hypothetical protein
	DQL34_RS00715	stage II sporulation protein M
	DQL34_RS05215	thioredoxin
	DQL34_RS08875	hypothetical protein
	DQL34_RS03970	RusA family crossover junction endodeoxyribonuclease
	DQL34_RS02015	hypothetical protein
	DQL34_RS05265	hypothetical protein
	DQL34_RS00695	hypothetical protein

	DQL34_RS05320	single-stranded DNA-binding protein
	DQL34_RS04005	hypothetical protein
	DQL34_RS03895	immunity protein
	DQL34_RS03995	helix-turn-helix domain-containing protein
	DQL34_RS04160	hypothetical protein
	DQL34_RS04120	DUF1366 domain-containing protein
	DQL34_RS02685	lactose transporter
	DQL34_RS04025	autolysin
	DQL34_RS09810	lactose transporter
	DQL34_RS05210	hypothetical protein
	DQL34_RS05330	hypothetical protein
	DQL34_RS07705	type II toxin-antitoxin system death-on-curing family toxin
	DQL34_RS04080	DUF3168 domain-containing protein
	DQL34_RS03195	hypothetical protein
	DQL34_RS05400	replication initiator protein A
	DQL34_RS02755	plasmid mobilization relaxosome protein MobC
	DQL34_RS07495	hypothetical protein
	DQL34_RS04055	hypothetical protein
	DQL34_RS04090	hypothetical protein
	DQL34_RS05225	hypothetical protein
	DQL34_RS03905	XRE family transcriptional regulator
	DQL34_RS08865	ribonuclease
	DQL34_RS05255	hypothetical protein
	DQL34_RS05405	hypothetical protein
	DQL34_RS05300	hypothetical protein
	DQL34_RS08870	hypothetical protein
	DQL34_RS04065	hypothetical protein
	DQL34_RS04075	histone H1
	DQL34_RS02665	hypothetical protein
	DQL34_RS02735	hypothetical protein
	DQL34_RS10500	transcriptional regulator
	DQL34_RS05250	hypothetical protein
	DQL34_RS02670	hypothetical protein
	DQL34_RS04140	hypothetical protein
	DQL34_RS04125	hypothetical protein

	DQL34_RS04095	hypothetical protein
	DQL34_RS05185	hypothetical protein
	DQL34_RS04070	hypothetical protein
	DQL34_RS03515	hypothetical protein
	DQL34_RS04000	DUF1372 family protein
	DQL34_RS03930	hypothetical protein
	DQL34_RS05360	hypothetical protein
	DQL34_RS05355	hypothetical protein
	DQL34_RS04540	hypothetical protein
	DQL34_RS00710	branched-chain amino acid aminotransferase
	DQL34_RS01330	hypothetical protein
	DQL34_RS05395	hypothetical protein
	DQL34_RS03925	excisionase
	DQL34_RS05245	hypothetical protein
	DQL34_RS03935	hypothetical protein
	DQL34_RS02710	hypothetical protein
	DQL34_RS07695	hypothetical protein
	DQL34_RS09745	transcriptional regulator
	DQL34_RS09785	hypothetical protein
	DQL34_RS05375	hypothetical protein
	DQL34_RS05275	hypothetical protein
	DQL34_RS05335	hypothetical protein
	DQL34_RS04150	hypothetical protein
	DQL34_RS04010	hypothetical protein
	DQL34_RS04130	phage holin
	DQL34_RS02780	hypothetical protein
	DQL34_RS03975	hypothetical protein
	DQL34_RS05220	hypothetical protein
	DQL34_RS10480	helix-turn-helix domain-containing protein
	DQL34_RS02675	hypothetical protein
	DQL34_RS03910	XRE family transcriptional regulator
	DQL34_RS08720	hypothetical protein
	DQL34_RS10790	DUF4649 family protein
	DQL34_RS10025	hypothetical protein
	DQL34_RS09805	hypothetical protein

	DQL34_RS09720	hypothetical protein
	DQL34_RS05205	hypothetical protein
	DQL34_RS10470	hypothetical protein
	DQL34_RS10235	hypothetical protein
	DQL34_RS02655	hypothetical protein
	DQL34_RS05175	hypothetical protein
	DQL34_RS05195	hypothetical protein
	DQL34_RS05285	hypothetical protein
	DQL34_RS08725	hypothetical protein
	DQL34_RS03980	hypothetical protein
	DQL34_RS03950	hypothetical protein
	DQL34_RS08885	hypothetical protein
	DQL34_RS04155	hypothetical protein
	DQL34_RS03285	type A2 lantipeptide
	DQL34_RS00315	transcriptional regulator
	DQL34_RS04530	DUF2974 domain-containing protein
	DQL34_RS01160	hypothetical protein
	DQL34_RS09620	bacteriocin
	DQL34_RS06245	glycosyltransferase
	DQL34_RS07500	hypothetical protein
	DQL34_RS04015	hypothetical protein

Supplementary Table S3. Genes involved in lactose and galactose metabolism identified in the 23 *S. thermophilus* strains. Genes highlighted in red correspond to putative pseudogenes

product	gene	ACA-DC 2	APC151	ASCC 1275	B59671	CNRZ1066
beta-galactosidase	*lacZ*	STACADC2_RS06290	B1761_RS02525	T303_RS07735	CG712_RS03095	STR_RS06575
lactose permease	*lacS*	STACADC2_RS06295	B1761_RS02530	T303_RS07740	CG712_RS03100	STR_RS06580
galactose 1-epimerase	*galM*	STACADC2_RS06300	B1761_RS02535	T303_RS07745	CG712_RS03105	STR_RS06585
UDP-glucose 4-epimerase	*galE*	STACADC2_RS06305	B1761_RS02540	T303_RS07750	CG712_RS03110	STR_RS06590
UDP-glucose--hexose-1-phosphate uridylyltransferase	*galT*	STACADC2_RS06310	B1761_RS02545	T303_RS07755	CG712_RS03115	STR_RS06595
galactokinase	*galK*	STACADC2_RS06315	B1761_RS02550	T303_RS07760	CG712_RS03120	STR_RS06600
transcriptional regulator	*galR*	STACADC2_RS06320	B1761_RS02555	T303_RS07765	CG712_RS03125	STR_RS06605
UDP-glucose 4-epimerase	*galE*	STACADC2_RS05320	B1761_RS01335	T303_RS06595	CG712_RS02060	STR_RS05510
UDP-glucose 4-epimerase	*galE*	-	-	-	-	-
UDP-glucose 4-epimerase	*galE*	-	-	-	-	-

product	gene	CS8	DGCC 7710	EPS	GABA	JIM 8232
beta-galactosidase	*lacZ*	BAY21_RS07785	CW339_RS06745	CR922_RS02815	CR921_RS06805	STH8232_RS06990
lactose permease	*lacS*	BAY21_RS07780	CW339_RS06750	CR922_RS02810	CR921_RS06810	STH8232_RS06995
galactose 1-epimerase	*galM*	BAY21_RS07775	CW339_RS06755	CR922_RS02805	CR921_RS06815	STH8232_RS07000
UDP-glucose 4-epimerase	*galE*	BAY21_RS07770	CW339_RS06760	CR922_RS02800	CR921_RS06820	STH8232_RS07005
UDP-glucose--hexose-1-phosphate uridylyltransferase	*galT*	BAY21_RS07765	CW339_RS06765	CR922_RS02795	CR921_RS06825	STH8232_RS07010
galactokinase	*galK*	BAY21_RS07760	CW339_RS06770	CR922_RS02790	CR921_RS06830	STH8232_RS07015
transcriptional regulator	*galR*	BAY21_RS07755	CW339_RS06775	CR922_RS02785	CR921_RS06835	STH8232_RS07020
UDP-glucose 4-epimerase	*galE*	BAY21_RS08850	CW339_RS05590	CR922_RS03905	CR921_RS05570	STH8232_RS05855
UDP-glucose 4-epimerase	*galE*	-	CW339_RS06575	-	-	-
UDP-glucose 4-epimerase	*galE*	-	-	-	CR921_RS06640	-

product	gene	KLDS 3.1003	KLDS SM	LMD-9	LMG 18311	MN-BM-A01
beta-galactosidase	*lacZ*	BEN15_RS09985	A9497_RS02955	STER_RS06725	STU_RS15920	AMD33_RS05650
lactose permease	*lacS*	BEN15_RS00005	A9497_RS02960	STER_RS06730	STU_RS15925	AMD33_RS05655
galactose 1-epimerase	*galM*	BEN15_RS00010	A9497_RS02965	STER_RS06735	STU_RS15930	AMD33_RS05660
UDP-glucose 4-epimerase	*galE*	BEN15_RS00015	A9497_RS02970	STER_RS06740	STU_RS15935	AMD33_RS05665
UDP-glucose--hexose-1-phosphate uridylyltransferase	*galT*	BEN15_RS00020	A9497_RS02975	STER_RS06745	STU_RS15940	AMD33_RS05670
galactokinase	*galK*	BEN15_RS00025	A9497_RS02980	STER_RS06750	STU_RS15945	AMD33_RS05675
transcriptional regulator	*galR*	BEN15_RS00030	A9497_RS02985	STER_RS06755	STU_RS15950	AMD33_RS05680

UDP-glucose 4-epimerase	*galE*	BEN15_RS08835	A9497_RS01825	STER_RS05565	STU_RS14900	AMD33_RS02935
UDP-glucose 4-epimerase	*galE*	-	-	-	-	-
UDP-glucose 4-epimerase	*galE*	-	-	-	-	-

product	**gene**	**MN-BM-A02**	**MN-ZLW-002**	**NCTC12958**	**ND03**	**ND07**
beta-galactosidase	*lacZ*	MNA02_RS06640	Y1U_RS06660	DQL34_RS07750	STND_RS06620	BBD27_RS07385
lactose permease	*lacS*	MNA02_RS06645	Y1U_RS06665	DQL34_RS07755	STND_RS06625	BBD27_RS07380
galactose 1-epimerase	*galM*	MNA02_RS06650	Y1U_RS06670	DQL34_RS07760	STND_RS06630	BBD27_RS07375
UDP-glucose 4-epimerase	*galE*	MNA02_RS06655	Y1U_RS06675	DQL34_RS07765	STND_RS06635	BBD27_RS07370
UDP-glucose--hexose-1-phosphate uridylyltransferase	*galT*	MNA02_RS06660	Y1U_RS06680	DQL34_RS07770	STND_RS06640	BBD27_RS07365
galactokinase	*galK*	MNA02_RS06665	Y1U_RS06685	DQL34_RS07775	STND_RS06645	BBD27_RS07360
transcriptional regulator	*galR*	MNA02_RS06670	Y1U_RS06690	DQL34_RS07780	STND_RS06650	BBD27_RS07355
UDP-glucose 4-epimerase	*galE*	MNA02_RS05500	Y1U_RS03970	DQL34_RS06535	STND_RS05450	BBD27_RS08525
UDP-glucose 4-epimerase	*galE*	-	-	-	-	-
UDP-glucose 4-epimerase	*galE*	-	-	DQL34_RS07605	-	-

product	**gene**	**S9**	**SMQ-301**	**ST3**
beta-galactosidase	*lacZ*	AVT04_RS07950	SMQ301_RS06710	BGL51_RS06915
lactose permease	*lacS*	AVT04_RS07945	SMQ301_RS06715	BGL51_RS06920
galactose 1-epimerase	*galM*	AVT04_RS07940	SMQ301_RS06720	BGL51_RS06925
UDP-glucose 4-epimerase	*galE*	AVT04_RS07935	SMQ301_RS06725	BGL51_RS06930
UDP-glucose--hexose-1-phosphate uridylyltransferase	*galT*	AVT04_RS07930	SMQ301_RS06730	BGL51_RS06935
galactokinase	*galK*	AVT04_RS07925	SMQ301_RS06735	BGL51_RS06940
transcriptional regulator	*galR*	AVT04_RS07920	SMQ301_RS06740	BGL51_RS06945
UDP-glucose 4-epimerase	*galE*	AVT04_RS09005	SMQ301_RS05560	BGL51_RS05785
UDP-glucose 4-epimerase	*galE*	-	-	-
UDP-glucose 4-epimerase	*galE*	-	SMQ301_RS06565	-

Supplementary Table S4. General characteristics of exopolysaccharide (EPS) clusters identified in the 23 *S. thermophilus* strains. Dashed lines are used to separate the two major clusters (A and B) and strain NCTC12958^T of the species, as described in the text

Strain	EPS coordinates	EPS length (bp)	Genes	CDS	Pseudogenes	Unique CDS	GC (%)
NCTC12958^T	com(1149594..1171095)	21,502	24	20	4	1	36.4
KLDS 3.1003	com(1605450..1627455)	22,006	24	15	9	1	34,9
ASCC 1275	com(1167275..1190619)	23,345	26	15	11	0	35,0
ND07	1663576..1686919	23,344	25	15	10	0	35,0
DGCC 7710	com(980395..1003738)	23,344	26	17	9	0	35,0
KLDS SM	com(292240..315583)	23,344	25	15	10	0	35,0
MN-BM-A02	com(977025..1000368)	23,344	25	15	10	0	35,0
MN-ZLW-002	792502..828474	35,973	41	32	9	1	36,0
MN-BM-A01	606167..642133	35,967	42	32	10	0	36,0
JIM 8232	com(1049941..1078711)	28,771	27	23	4	3	36,4
LMD-9	com(975018..996065)	21,048	21	16	5	5	34,7
SMQ-301	com(977441..1005388)	27,948	27	17	10	4	34,3
ND03	com(951312..977745)	26,434	34	23	11	0	35,1
APC151	com(163250..189683)	26,434	34	23	11	0	35,1
GABA	com(979718..1002037)	22,320	24	19	5	0	35.7
ST3	com(1000985..1025295)	24,311	28	20	8	0	34,6
CNRZ1066	com(960253..981421)	21,169	22	18	4	0	34,7
CS8	1725876..1747044	21,169	22	17	5	0	34,7
S9	1743015..1764183	21,169	22	17	5	0	34,7
EPS	783382..804549	21,168	23	18	5	0	34,7
LMG 18311	com(954247..983323)	29,077	36	27	9	15	34,3
B59671	com(316701..347524)	30,824	33	24	9	9	35,6
ACA-DC 2	com(928529..947189)	18,661	21	20	1	6	34,8

Supplementary Table S5A. Genes involved in proteolysis identified in the 23 *S. thermophilus* strains. Genes highlighted in blue are truncated

product	gene	ACA-DC 2	APC151	ASCC 1275	B59671	CNRZ1066
Aminopetidase A / Glutamyl aminopeptidase	*pepA*	STACADC2_RS08405	B1761_RS04700	T303_RS00190	CG712_RS05205	STR_RS08685
Aminopeptidase C	*pepC*	STACADC2_RS01215	B1761_RS06920	T303_RS02310	CG712_RS07285	STR_RS01315
Oligoendopeptidase F	*pepF*	STACADC2_RS02300	B1761_RS08050	T303_RS03355	CG712_RS08340	STR_RS02365
Methionine aminopeptidase	*pepM*	STACADC2_RS07005	B1761_RS03275	T303_RS08470	CG712_RS03795	STR_RS07300
Aminopeptidase N	*pepN*	STACADC2_RS04765	B1761_RS00725	T303_RS05990	CG712_RS01295	STR_RS04900
Endopeptidase	*pepO*	STACADC2_RS08575	B1761_RS04875	T303_RS00365	CG712_RS05370	STR_RS08855
Aminopeptidase P family protein	*pepP*	STACADC2_RS07875	B1761_RS04160	T303_RS09350	CG712_RS04675	STR_RS08155
Prolidase / Dipeptidase	*pepQ*	STACADC2_RS03170	B1761_RS08940	T303_RS04280	CG712_RS09210	STR_RS03225
Aminopeptidase PepS	*pepS*	STACADC2_RS00455	B1761_RS06085	T303_RS01565	CG712_RS06485	STR_RS00475
Peptidase T (tripeptidase)	*pepT*	STACADC2_RS05180	B1761_RS01195	T303_RS06455	CG712_RS01920	STR_RS05370
Dipeptidase PepV	*pepV*	STACADC2_RS05125	B1761_RS01140	T303_RS06405	CG712_RS01865	STR_RS05315
Xaa-Pro dipeptidyl-peptidase	*pepX*	STACADC2_RS07555	B1761_RS03840	T303_RS09015	CG712_RS04370	STR_RS07875
peptidase S51	-	-	-	-	-	-
cell-wall associated proteinase / serine protease	*prtS*	-	-	T303_RS05160	-	-

product	gene	CS8	DGCC 7710	EPS	GABA	JIM 8232
Aminopetidase A / Glutamyl aminopeptidase	*pepA*	BAY21_RS05670	CW339_RS08905	CR922_RS08845	CR921_RS08910	STH8232_RS09155
Aminopeptidase C	*pepC*	BAY21_RS03595	CW339_RS01180	CR922_RS08155	CR921_RS01255	STH8232_RS01425
Oligoendopeptidase F	*pepF*	BAY21_RS02540	CW339_RS02250	CR922_RS07085	CR921_RS02365	STH8232_RS02475
Methionine aminopeptidase	*pepM*	BAY21_RS07070	CW339_RS07475	CR922_RS02065	CR921_RS07560	STH8232_RS07695
Aminopeptidase N	*pepN*	BAY21_RS00035	CW339_RS04960	CR922_RS04525	CR921_RS04995	STH8232_RS05260
Endopeptidase	*pepO*	BAY21_RS05500	CW339_RS09080	CR922_RS09015	CR921_RS09085	STH8232_RS09330
Aminopeptidase P family protein	*pepP*	BAY21_RS06210	CW339_RS08370	CR922_RS01170	CR921_RS08385	STH8232_RS08615
Prolidase / Dipeptidase	*pepQ*	BAY21_RS01690	CW339_RS03215	CR922_RS06235	CR921_RS03320	STH8232_RS03605
Aminopeptidase PepS	*pepS*	BAY21_RS04365	CW339_RS00450	CR922_RS00455	CR921_RS00460	STH8232_RS00515
Peptidase T (tripeptidase)	*pepT*	BAY21_RS08990	CW339_RS05450	CR922_RS04045	CR921_RS05430	STH8232_RS05715
Dipeptidase PepV	*pepV*	BAY21_RS09040	CW339_RS05395	CR922_RS04100	CR921_RS05375	STH8232_RS05660
Xaa-Pro dipeptidyl-peptidase	*pepX*	BAY21_RS06495	CW339_RS08040	CR922_RS01465	CR921_RS08150	STH8232_RS08280
peptidase S51	-	-	-	-	-	STH8232_RS00675
cell-wall associated proteinase / serine protease	*prtS*	-	CW339_RS04105	-	CR921_RS04045	STH8232_RS04360

product	gene	KLDS 3.1003	KLDS SM	LMD-9	LMG 18311	MN-BM-A01
Aminopetidase A / Glutamyl aminopeptidase	*pepA*	BEN15_RS02180	A9497_RS05110	STER_RS08925	STU_RS18080	AMD33_RS07785
Aminopeptidase C	*pepC*	BEN15_RS04300	A9497_RS07295	STER_RS01340	STU_RS10780	AMD33_RS00200
Oligoendopeptidase F	*pepF*	BEN15_RS05450	A9497_RS08360	STER_RS02405	STU_RS11845	AMD33_RS01300
Methionine aminopeptidase	*pepM*	BEN15_RS00815	A9497_RS03680	STER_RS07460	STU_RS16650	AMD33_RS06370
Aminopeptidase N	*pepN*	BEN15_RS08185	A9497_RS01225	STER_RS04990	STU_RS14255	AMD33_RS03580
Endopeptidase	*pepO*	BEN15_RS02350	A9497_RS05285	STER_RS09100	STU_RS18250	AMD33_RS07955
Aminopeptidase P family protein	*pepP*	BEN15_RS01645	A9497_RS04560	STER_RS08390	STU_RS17535	AMD33_RS07235
Prolidase / Dipeptidase	*pepQ*	BEN15_RS06310	A9497_RS09300	STER_RS03340	STU_RS12700	AMD33_RS02155
Aminopeptidase PepS	*pepS*	BEN15_RS03530	A9497_RS06485	STER_RS00485	STU_RS09945	AMD33_RS09155
Peptidase T (tripeptidase)	*pepT*	BEN15_RS08675	A9497_RS01685	STER_RS05425	STU_RS14765	AMD33_RS03075
Dipeptidase PepV	*pepV*	BEN15_RS08625	A9497_RS01640	STER_RS05375	STU_RS14710	AMD33_RS03130
Xaa-Pro dipeptidyl-peptidase	*pepX*	BEN15_RS01405	A9497_RS04230	STER_RS08015	STU_RS17225	AMD33_RS06920
peptidase S51	-	-	-	-	-	-
cell-wall associated proteinase / serine protease	*prtS*	-	A9497_RS00420	STER_RS04165	-	AMD33_RS04405

product	gene	MN-BM-A02	MN-ZLW-002	NCTC12958	ND03	ND07
Aminopetidase A / Glutamyl aminopeptidase	*pepA*	MNA02_RS08805	Y1U_RS08800	DQL34_RS10270	STND_RS08790	BBD27_RS05235
Aminopeptidase C	*pepC*	MNA02_RS01220	Y1U_RS01245	DQL34_RS01565	STND_RS01280	BBD27_RS03130
Oligoendopeptidase F	*pepF*	MNA02_RS02275	Y1U_RS02340	DQL34_RS02640	STND_RS02395	BBD27_RS02075
Methionine aminopeptidase	*pepM*	MNA02_RS07375	Y1U_RS07390	DQL34_RS08520	STND_RS07380	BBD27_RS06670
Aminopeptidase N	*pepN*	MNA02_RS04900	Y1U_RS04595	DQL34_RS05965	STND_RS04845	BBD27_RS09125
Endopeptidase	*pepO*	MNA02_RS08980	Y1U_RS08965	DQL34_RS10420	STND_RS08965	BBD27_RS05060
Aminopeptidase P family protein	*pepP*	MNA02_RS08260	Y1U_RS08255	DQL34_RS09600	STND_RS08240	BBD27_RS05785
Prolidase / Dipeptidase	*pepQ*	MNA02_RS03200	Y1U_RS03200	DQL34_RS03855	STND_RS03275	BBD27_RS01130
Aminopeptidase PepS	*pepS*	MNA02_RS00475	Y1U_RS00480	DQL34_RS00530	STND_RS00515	BBD27_RS03865
Peptidase T (tripeptidase)	*pepT*	MNA02_RS05360	Y1U_RS04110	DQL34_RS06395	STND_RS05310	BBD27_RS08665
Dipeptidase PepV	*pepV*	MNA02_RS05310	Y1U_RS04165	DQL34_RS06340	STND_RS05255	BBD27_RS08715
Xaa-Pro dipeptidyl-peptidase	*pepX*	MNA02_RS07925	Y1U_RS07935	DQL34_RS09235	STND_RS07925	BBD27_RS06115
peptidase S51	-	-	-	-	-	-
cell-wall associated proteinase / serine protease	*prtS*	MNA02_RS04075	Y1U_RS05420	-	-	BBD27_RS00265

product	gene	S9	SMQ-301	ST3
Aminopetidase A / Glutamyl aminopeptidase	*pepA*	AVT04_RS05825	SMQ301_RS08915	BGL51_RS09005
Aminopeptidase C	*pepC*	AVT04_RS03675	SMQ301_RS01355	BGL51_RS01300
Oligoendopeptidase F	*pepF*	AVT04_RS02620	SMQ301_RS02415	BGL51_RS02360
Methionine aminopeptidase	*pepM*	AVT04_RS07230	SMQ301_RS07445	BGL51_RS07665
Aminopeptidase N	*pepN*	AVT04_RS00145	SMQ301_RS04970	BGL51_RS05195
Endopeptidase	*pepO*	AVT04_RS05655	SMQ301_RS09085	BGL51_RS09180
Aminopeptidase P family protein	*pepP*	AVT04_RS06365	SMQ301_RS08380	BGL51_RS08495
Prolidase / Dipeptidase	*pepQ*	AVT04_RS01770	SMQ301_RS03320	BGL51_RS03305
Aminopeptidase PepS	*pepS*	AVT04_RS04525	SMQ301_RS00490	BGL51_RS00455
Peptidase T (tripeptidase)	*pepT*	AVT04_RS09145	SMQ301_RS05420	BGL51_RS05640
Dipeptidase PepV	*pepV*	AVT04_RS09200	SMQ301_RS05370	BGL51_RS05585
Xaa-Pro dipeptidyl-peptidase	*pepX*	AVT04_RS06655	SMQ301_RS08005	BGL51_RS08255
peptidase S51	-	-	-	-
cell-wall associated proteinase / serine protease	*prtS*	-	SMQ301_RS04180	BGL51_RS04150

Supplamentary Table S5B. Genes involved in transport of peptides and amino acids identified in the 23 *S. thermophilus* strains. Genes highlighted in red correspond to putative pseudogenes while genes highlighted in blue are truncated

Transporter family	product		gene	ACA-DC 2	APC151	ASCC 1275	B59671	CNRZ1066
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	Oligopeptide transport ATP-binding protein OppF	*OppF*	STACADC2_RS06495	B1761_RS02725	T303_RS07925	CG712_RS03285	STR_RS06760
		Oligopeptide transport ATP-binding protein OppD	*OppD*	STACADC2_RS06500	B1761_RS02730	T303_RS07930	CG712_RS03290	STR_RS06765
		Oligopeptide transport system permease protein OppC	*OppC*	STACADC2_RS06505	B1761_RS02735	T303_RS07935	CG712_RS03295	STR_RS06770
		Oligopeptide transport system permease protein OppB	*OppB*	STACADC2_RS06510	B1761_RS02740	T303_RS07940	CG712_RS03300	STR_RS06775
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	STACADC2_RS06515	B1761_RS02745	T303_RS07945	CG712_RS03305	STR_RS06780
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	STACADC2_RS06525	B1761_RS02755	T303_RS07955	CG712_RS06835	STR_RS06790
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	peptide ABC transporter substrate-binding protein	*OppA*	-	-	-	-	-
		ABC transporter permease	*OppB*	-	-	-	-	-
		ABC transporter permease	*OppC*	-	-	-	-	-
		ABC transporter ATP-binding protein	*OppD*	-	-	-	-	-
ATP-binding Cassette (ABC)	branched-chain amino acids ABC transporter	branched-chain amino acid ABC transporter substrate-binding protein	*livK*	STACADC2_RS01820	B1761_RS07535	T303_RS02910	CG712_RS07855	STR_RS01880
		branched-chain amino acid ABC transporter permease	*livH*	STACADC2_RS01825	B1761_RS07540	T303_RS02915	CG712_RS07860	STR_RS01885
		branched-chain amino acid ABC transporter permease	*livM*	STACADC2_RS01830	B1761_RS07545	T303_RS02920	CG712_RS07865	STR_RS01890
		ABC transporter ATP-binding protein	*livG*	STACADC2_RS01835	B1761_RS07550	T303_RS02925	CG712_RS07870	STR_RS01895
		ABC transporter ATP-binding protein	*livF*	STACADC2_RS01840	B1761_RS07555	T303_RS02930	CG712_RS07875	STR_RS01900

ATP-binding Cassette (ABC)	glutamine ABC transporter	amino acid ABC transporter permease	-	STACADC2_RS06745	B1761_RS03005	T303_RS08195	CG712_RS03540	STR_RS07025
		amino acid ABC transporter ATP-binding protein	-	STACADC2_RS06750	B1761_RS03010	T303_RS08200	CG712_RS03545	STR_RS07030
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter permease	-	STACADC2_RS04190	B1761_RS00130	T303_RS05485	CG712_RS00755	STR_RS04350
		amino acid ABC transporter ATP-binding protein	-	STACADC2_RS04195	B1761_RS00135	T303_RS05490	CG712_RS00760	STR_RS04355
		amino acid ABC transporter substrate-binding protein	-	STACADC2_RS04200	B1761_RS00140	T303_RS05495	CG712_RS00765	STR_RS04360
ATP-binding Cassette (ABC)	polar amino acid (cysteine) ABC transporter	amino acid ABC transporter substrate-binding protein	-	STACADC2_RS07115	B1761_RS03385	T303_RS08575	CG712_RS03905	STR_RS07410
		amino acid ABC transporter ATP-binding protein	-	STACADC2_RS07120	B1761_RS03390	T303_RS08580	CG712_RS03910	STR_RS07415
		amino acid ABC transporter permease	-	STACADC2_RS07125	B1761_RS03395	T303_RS08585	CG712_RS03915	STR_RS07420
		amino acid ABC transporter permease	-	STACADC2_RS07130	B1761_RS03400	T303_RS08590	CG712_RS03920	STR_RS07425
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	STACADC2_RS07485	B1761_RS03755	T303_RS08935	CG712_RS04275	STR_RS07770
		amino acid ABC transporter permease	-	STACADC2_RS07490	B1761_RS03760	T303_RS08940	CG712_RS04280	STR_RS07775
		amino acid ABC transporter substrate-binding protein	-	STACADC2_RS07495	B1761_RS03765	T303_RS08945	CG712_RS04285	STR_RS07780
ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	STACADC2_RS05280	B1761_RS01295	T303_RS06555	CG712_RS02020	STR_RS05470
		glutamine ABC transporter substrate-binding protein	-	STACADC2_RS05285	B1761_RS01300	T303_RS06560	CG712_RS02025	STR_RS05475
		amino acid ABC transporter permease	-	STACADC2_RS05290	B1761_RS01305	T303_RS06565	CG712_RS02030	STR_RS05480
		amino acid ABC transporter permease	-	STACADC2_RS05295	B1761_RS01310	T303_RS06570	CG712_RS02035	STR_RS05485
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	-	-	-	CG712_RS02915	-
		amino acid ABC transporter permease	-	-	-	-	CG712_RS02920	-
		amino acid ABC transporter permease	-	-	-	-	CG712_RS02925	-
		amino acid ABC transporter substrate-binding protein	-	-	-	-	CG712_RS02930	-

ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter substrate-binding protein	-	-	-	-	-	STR_RS06350
		amino acid ABC transporter ATP-binding protein	-	-	-	-	-	STR_RS06360
		amino acid ABC transporter permease	-	-	-	-	-	STR_RS06365
ATP-binding Cassette (ABC)	spermidine/putrescine ABC transporter	spermidine/putrescine ABC transporter, periplasmic binding protein PotD	*potD*	STACADC2_RS06890	B1761_RS03165	T303_RS08355	CG712_RS03680	STR_RS07185
		spermidine/putrescine ABC transporter, permease component potC	*potC*	STACADC2_RS06895	B1761_RS03170	T303_RS08360	CG712_RS03685	STR_RS07190
		spermidine/putrescine ABC transporter, permease component PotB	*potB*	STACADC2_RS06900	B1761_RS03175	T303_RS08365	CG712_RS03690	STR_RS07195
		Putrescine transport ATP-binding protein PotA	*potA*	STACADC2_RS06905	B1761_RS03180	T303_RS08370	CG712_RS03695	STR_RS07200
ATP-binding Cassette (ABC)	methionine ABC transporter	amino acid ABC transporter substrate protein	*metQ*	STACADC2_RS01535	B1761_RS07255	T303_RS02630	CG712_RS07575	STR_RS01610
		methionine import ATP-binding protein MetN	*metN*	STACADC2_RS01545	B1761_RS07265	T303_RS02640	CG712_RS07585	STR_RS01620
		methionine ABC transporter permease	*metP*	STACADC2_RS01550	B1761_RS07270	T303_RS02645	CG712_RS07590	STR_RS01625
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (asparagine)	-	STACADC2_RS02555	B1761_RS08305	T303_RS03625	CG712_RS08600	STR_RS02620
Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	STACADC2_RS02900	B1761_RS08680	T303_RS04035	CG712_RS08960	STR_RS02975
Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	STACADC2_RS06145	B1761_RS02320	T303_RS07535	CG712_RS02895	STR_RS06425
Amino Acid-Polyamine-Organocation (APC)		amino acid specific permease	-	STACADC2_RS06940	B1761_RS03215	T303_RS08405	CG712_RS03735	STR_RS07235
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (glutamate:GABA antiporter)	*gadC*	STACADC2_RS02050	B1761_RS07770	-	CG712_RS08085	-
Amino Acid-Polyamine-Organocation (APC)		histidine:histamine antiporter	*hdcP*	-	-	-	-	-
Branched Chain Amino Acid Exporter (LIV-E)		branched-chain amino acid permease (AzlC)	-	STACADC2_RS08005	B1761_RS04290	T303_RS09480	CG712_RS04805	STR_RS08290

Resistance to Homoserine/Threonine (RhtB)		LysE family translocator (amino acid efflux protein)	-	STACADC2_RS06245	B1761_RS02425	T303_RS07645	-	STR_RS06540
Proton-dependent Oligopeptide Transporter (POT)	di-tripeptide transporter	peptide MFS transporter	*dtpT*	STACADC2_RS04610	B1761_RS00560	T303_RS05850	CG712_RS01150	STR_RS04745
Branched Chain Amino Acid:Cation Symporter (LIVCS)	transporter (symporter)	branched-chain amino acid transport system II carrier protein	*brnQ*	STACADC2_RS06120	B1761_RS02290	T303_RS07515	CG712_RS02870	STR_RS06405
Alanine or Glycine:Cation Symporter (AGCS)	transporter (symporter)	sodium:alanine symporter family protein	-	STACADC2_RS04660	B1761_RS00620	T303_RS05890	CG712_RS01190	STR_RS04800
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		Na+/serine-threonine symporter	*sstT*	STACADC2_RS01555	B1761_RS07275	T303_RS02650	CG712_RS07595	STR_RS01630
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		dicarboxylate/amino acid:cation symporter	-	-	B1761_RS04925	-	-	-

Transporter family	**product**		**gene**	**CS8**	**DGCC 7710**	**EPS**	**GABA**	**JIM 8232**
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	Oligopeptide transport ATP-binding protein OppF	*OppF*	BAY21_RS07600	CW339_RS06950	CR922_RS02625	CR921_RS07000	STH8232_RS07180
		Oligopeptide transport ATP-binding protein OppD	*OppD*	BAY21_RS07595	CW339_RS06955	CR922_RS02620	CR921_RS07005	STH8232_RS07185
		Oligopeptide transport system permease protein OppC	*OppC*	BAY21_RS07590	CW339_RS06960	CR922_RS02615	CR921_RS07010	STH8232_RS07190
		Oligopeptide transport system permease protein OppB	*OppB*	BAY21_RS07585	CW339_RS06965	CR922_RS02610	CR921_RS07015	STH8232_RS07195
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	BAY21_RS07580	CW339_RS06970	CR922_RS02605	CR921_RS07020	STH8232_RS07200
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	BAY21_RS07575	CW339_RS06980	CR922_RS02595	CR921_RS07030	STH8232_RS07210
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	peptide ABC transporter substrate-binding protein	*OppA*	-	-	-	-	-
		ABC transporter permease	*OppB*	-	-	-	-	-

		ABC transporter permease	*OppC*	-	-	-	-	-
		ABC transporter ATP-binding protein	*OppD*	-	-	-	-	-
ATP-binding Cassette (ABC)	branched-chain amino acids ABC transporter	branched-chain amino acid ABC transporter substrate-binding protein	*livK*	BAY21_RS03020	CW339_RS01795	CR922_RS07565	CR921_RS01860	STH8232_RS02020
		branched-chain amino acid ABC transporter permease	*livH*	BAY21_RS03015	CW339_RS01800	CR922_RS07560	CR921_RS01865	STH8232_RS02025
		branched-chain amino acid ABC transporter permease	*livM*	BAY21_RS03010	CW339_RS01805	CR922_RS07555	CR921_RS01870	STH8232_RS02030
		ABC transporter ATP-binding protein	*livG*	BAY21_RS03005	CW339_RS01810	CR922_RS07550	CR921_RS01875	STH8232_RS02035
		ABC transporter ATP-binding protein	*livF*	BAY21_RS03000	CW339_RS01815	CR922_RS07545	CR921_RS01880	STH8232_RS02040
ATP-binding Cassette (ABC)	glutamine ABC transporter	amino acid ABC transporter permease	-	BAY21_RS07345	CW339_RS07205	CR922_RS02360	CR921_RS07285	STH8232_RS07440
		amino acid ABC transporter ATP-binding protein	-	BAY21_RS07340	CW339_RS07210	CR922_RS02355	CR921_RS07290	STH8232_RS07445
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter permease	-	BAY21_RS00585	CW339_RS04425	CR922_RS05090	CR921_RS04370	STH8232_RS04670
		amino acid ABC transporter ATP-binding protein	-	BAY21_RS00580	CW339_RS04430	CR922_RS05085	CR921_RS04375	STH8232_RS04675
		amino acid ABC transporter substrate-binding protein	-	BAY21_RS00575	CW339_RS04435	CR922_RS05080	CR921_RS04380	STH8232_RS04680
ATP-binding Cassette (ABC)	polar amino acid (cysteine) ABC transporter	amino acid ABC transporter substrate-binding protein	-	BAY21_RS06960	CW339_RS07580	CR922_RS01955	CR921_RS07665	STH8232_RS07840
		amino acid ABC transporter ATP-binding protein	-	BAY21_RS06955	CW339_RS07585	CR922_RS01950	CR921_RS07670	STH8232_RS07845
		amino acid ABC transporter permease	-	BAY21_RS06950	CW339_RS07590	CR922_RS01945	CR921_RS07675	STH8232_RS07850
		amino acid ABC transporter permease	-	BAY21_RS06945	CW339_RS07595	CR922_RS01940	CR921_RS07680	STH8232_RS07855
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	BAY21_RS06600	CW339_RS07955	CR922_RS01580	CR921_RS08060	STH8232_RS08200
		amino acid ABC transporter permease	-	BAY21_RS06595	CW339_RS07960	CR922_RS01575	CR921_RS08065	STH8232_RS08205
		amino acid ABC transporter substrate-binding protein	-	BAY21_RS06590	CW339_RS07965	CR922_RS01570	CR921_RS08070	STH8232_RS08210
ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	BAY21_RS08890	CW339_RS05550	CR922_RS03945	CR921_RS05530	STH8232_RS05815

		glutamine ABC transporter substrate-binding protein	-	BAY21_RS08885	CW339_RS05555	CR922_RS03940	CR921_RS05535	STH8232_RS05820
		amino acid ABC transporter permease	-	BAY21_RS08880	CW339_RS05560	CR922_RS03935	CR921_RS05540	STH8232_RS05825
		amino acid ABC transporter permease	-	BAY21_RS08875	CW339_RS05565	CR922_RS03930	CR921_RS05545	STH8232_RS05830
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	-	-	-	CR921_RS06545	-
		amino acid ABC transporter permease	-	-	-	-	CR921_RS06550	-
		amino acid ABC transporter permease	-	-	-	-	CR921_RS06555	-
		amino acid ABC transporter substrate-binding protein	-	-	-	-	CR921_RS06560	-
ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter substrate-binding protein	-	BAY21_RS08010	-	CR922_RS03050	-	-
		amino acid ABC transporter ATP-binding protein	-	BAY21_RS08000	-	CR922_RS03040	-	-
		amino acid ABC transporter permease	-	BAY21_RS07995	-	CR922_RS03035	-	-
ATP-binding Cassette (ABC)	spermidine/putrescine ABC transporter	spermidine/putrescine ABC transporter, periplasmic binding protein PotD	*potD*	BAY21_RS07180	CW339_RS07360	CR922_RS02180	CR921_RS07440	STH8232_RS07580
		spermidine/putrescine ABC transporter, permease component potC	*potC*	BAY21_RS07175	CW339_RS07365	CR922_RS02175	CR921_RS07445	STH8232_RS07585
		spermidine/putrescine ABC transporter, permease component PotB	*potB*	BAY21_RS07170	CW339_RS07370	CR922_RS02170	CR921_RS07450	STH8232_RS07590
		Putrescine transport ATP-binding protein PotA	*potA*	BAY21_RS07165	CW339_RS07375	CR922_RS02165	CR921_RS07455	STH8232_RS07595
ATP-binding Cassette (ABC)	methionine ABC transporter	amino acid ABC transporter substrate protein	*metQ*	BAY21_RS03290	CW339_RS01515	-	CR921_RS01575	STH8232_RS01740
		methionine import ATP-binding protein MetN	*metN*	BAY21_RS03280	CW339_RS01525	CR922_RS07835	CR921_RS01585	STH8232_RS01750
		methionine ABC transporter permease	*metP*	BAY21_RS03275	CW339_RS01530	CR922_RS07830	CR921_RS01590	STH8232_RS01755
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (asparagine)	-	BAY21_RS02285	CW339_RS02515	CR922_RS06835	CR921_RS02620	STH8232_RS02750

Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	BAY21_RS01935	CW339_RS02955	CR922_RS06480	CR921_RS03075	STH8232_RS03370
Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	BAY21_RS07935	CW339_RS06530	CR922_RS02970	CR921_RS06525	STH8232_RS06785
Amino Acid-Polyamine-Organocation (APC)		amino acid specific permease	-	BAY21_RS07130	CW339_RS07410	CR922_RS02130	CR921_RS07490	STH8232_RS07630
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (glutamate:GABA antiporter)	*gadC*	-	-	-	CR921_RS02095	-
Amino Acid-Polyamine-Organocation (APC)		histidine:histamine antiporter	*hdcP*	-	-	-	-	-
Branched Chain Amino Acid Exporter (LIV-E)		branched-chain amino acid permease (AzlC)	-	BAY21_RS06075	CW339_RS08500	CR922_RS01040	CR921_RS08515	STH8232_RS08745
Resistance to Homoserine/Threonine (RhtB)		LysE family translocator (amino acid efflux protein)	-	BAY21_RS07820	CW339_RS06640	CR922_RS02860	CR921_RS06700	STH8232_RS06895
Proton-dependent Oligopeptide Transporter (POT)	di-tripeptide transporter	peptide MFS transporter	*dtpT*	BAY21_RS00190	CW339_RS04815	CR922_RS04685	CR921_RS04850	STH8232_RS05120
Branched Chain Amino Acid:Cation Symporter (LIVCS)	transporter (symporter)	branched-chain amino acid transport system II carrier protein	*brnQ*	BAY21_RS07955	CW339_RS06505	CR922_RS02995	CR921_RS06500	STH8232_RS06765
Alanine or Glycine:Cation Symporter (AGCS)	transporter (symporter)	sodium:alanine symporter family protein	-	BAY21_RS00135	CW339_RS04855	CR922_RS04630	CR921_RS04890	STH8232_RS05160
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		Na+/serine-threonine symporter	*sstT*	BAY21_RS03270	CW339_RS01535	CR922_RS07825	CR921_RS01595	STH8232_RS01760
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		dicarboxylate/amino acid:cation symporter	-	-	-	-	-	-

Transporter family	**product**		**gene**	**KLDS 3.1003**	**KLDS SM**	**LMD-9**	**LMG 18311**	**MN-BM-A01**
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	Oligopeptide transport ATP-binding protein OppF	*OppF*	BEN15_RS00250	A9497_RS03140	STER_RS06910	STU_RS16110	AMD33_RS05835
		Oligopeptide transport ATP-binding protein OppD	*OppD*	BEN15_RS00255	A9497_RS03145	STER_RS06915	STU_RS16115	AMD33_RS05840
		Oligopeptide transport system permease protein OppC	*OppC*	BEN15_RS00260	A9497_RS03150	STER_RS06920	STU_RS16120	AMD33_RS05845

		Oligopeptide transport system permease protein OppB	*OppB*	BEN15_RS00265	A9497_RS03155	STER_RS06925	STU_RS16125	AMD33_RS05850
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	BEN15_RS00270	A9497_RS03160	STER_RS06930	STU_RS16130	AMD33_RS05855
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	BEN15_RS00280	A9497_RS03170	STER_RS06940	STU_RS16140	AMD33_RS05865
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	peptide ABC transporter substrate-binding protein	*OppA*	-	-	STER_RS00685	-	-
		ABC transporter permease	*OppB*	-	-	STER_RS00690	-	-
		ABC transporter permease	*OppC*	-	-	STER_RS00695	-	-
		ABC transporter ATP-binding protein	*OppD*	-	-	STER_RS00700	-	-
ATP-binding Cassette (ABC)	branched-chain amino acids ABC transporter	branched-chain amino acid ABC transporter substrate-binding protein	*livK*	BEN15_RS04920	A9497_RS07905	STER_RS01940	STU_RS11380	AMD33_RS00800
		branched-chain amino acid ABC transporter permease	*livH*	BEN15_RS04925	A9497_RS07910	STER_RS01945	STU_RS11385	AMD33_RS00805
		branched-chain amino acid ABC transporter permease	*livM*	BEN15_RS04930	A9497_RS07915	STER_RS01950	STU_RS11390	AMD33_RS00810
		ABC transporter ATP-binding protein	*livG*	BEN15_RS04935	A9497_RS07920	STER_RS01955	STU_RS11395	AMD33_RS00815
		ABC transporter ATP-binding protein	*livF*	BEN15_RS04940	A9497_RS07925	STER_RS01960	STU_RS11400	AMD33_RS00820
ATP-binding Cassette (ABC)	glutamine ABC transporter	amino acid ABC transporter permease	-	BEN15_RS00540	A9497_RS03405	STER_RS07190	STU_RS16365	AMD33_RS06095
		amino acid ABC transporter ATP-binding protein	-	BEN15_RS00545	A9497_RS03410	STER_RS07195	STU_RS16370	AMD33_RS06100
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter permease	-	BEN15_RS07610	A9497_RS00720	STER_RS04480	STU_RS13695	AMD33_RS04100
		amino acid ABC transporter ATP-binding protein	-	BEN15_RS07615	A9497_RS00725	STER_RS04485	STU_RS13700	AMD33_RS04095
		amino acid ABC transporter substrate-binding protein	-	BEN15_RS07620	A9497_RS00730	STER_RS04490	STU_RS13705	AMD33_RS04090
ATP-binding Cassette (ABC)	polar amino acid (cysteine) ABC transporter	amino acid ABC transporter substrate-binding protein	-	BEN15_RS00910	A9497_RS03785	STER_RS07565	STU_RS16760	AMD33_RS06475

		amino acid ABC transporter ATP-binding protein	-	BEN15_RS00915	A9497_RS03790	STER_RS07570	STU_RS16765	AMD33_RS06480
		amino acid ABC transporter permease	-	BEN15_RS00920	A9497_RS03795	STER_RS07575	STU_RS16770	AMD33_RS06485
		amino acid ABC transporter permease	-	BEN15_RS00925	A9497_RS03800	STER_RS07580	STU_RS16775	AMD33_RS06490
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	BEN15_RS01310	A9497_RS04145	STER_RS07935	STU_RS17120	AMD33_RS06835
		amino acid ABC transporter permease	-	BEN15_RS01315	A9497_RS04150	STER_RS07940	STU_RS17125	AMD33_RS06840
		amino acid ABC transporter substrate-binding protein	-	BEN15_RS01320	A9497_RS04155	STER_RS07945	STU_RS17130	AMD33_RS06845
ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	BEN15_RS08795	A9497_RS01785	STER_RS05525	STU_RS14860	AMD33_RS02975
		glutamine ABC transporter substrate-binding protein	-	BEN15_RS08800	A9497_RS01790	STER_RS05530	STU_RS14865	AMD33_RS02970
		amino acid ABC transporter permease	-	BEN15_RS08805	A9497_RS01795	STER_RS05535	STU_RS14870	AMD33_RS02965
		amino acid ABC transporter permease	-	BEN15_RS08810	A9497_RS01800	STER_RS05540	STU_RS14875	AMD33_RS02960
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	-	-	-	-	-
		amino acid ABC transporter permease	-	-	-	-	-	-
		amino acid ABC transporter permease	-	-	-	-	-	-
		amino acid ABC transporter substrate-binding protein	-	-	-	-	-	-
ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter substrate-binding protein	-	BEN15_RS09750	-	-	-	-
		amino acid ABC transporter ATP-binding protein	-	BEN15_RS09760	-	-	-	-
		amino acid ABC transporter permease	-	BEN15_RS09765	-	-	-	-
ATP-binding Cassette (ABC)	spermidine/putrescine ABC transporter	spermidine/putrescine ABC transporter, periplasmic binding protein PotD	*potD*	BEN15_RS00700	A9497_RS03570	STER_RS07345	STU_RS16535	AMD33_RS06255
		spermidine/putrescine ABC transporter, permease component potC	*potC*	BEN15_RS00705	A9497_RS03575	STER_RS07350	STU_RS16540	AMD33_RS06260

		spermidine/putrescine ABC transporter, permease component PotB	*potB*	BEN15_RS00710	A9497_RS03580	STER_RS07355	STU_RS16545	AMD33_RS06265
		Putrescine transport ATP-binding protein PotA	*potA*	BEN15_RS00715	A9497_RS03585	STER_RS07360	STU_RS16550	AMD33_RS06270
ATP-binding Cassette (ABC)	methionine ABC transporter	amino acid ABC transporter substrate protein	*metQ*	BEN15_RS04625	A9497_RS07625	STER_RS01655	STU_RS11105	AMD33_RS00530
		methionine import ATP-binding protein MetN	*metN*	BEN15_RS04635	A9497_RS07635	STER_RS01665	STU_RS11115	AMD33_RS00540
		methionine ABC transporter permease	*metP*	BEN15_RS04640	A9497_RS07640	STER_RS01670	STU_RS11120	AMD33_RS00545
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (asparagine)	-	BEN15_RS05715	A9497_RS08630	STER_RS02670	STU_RS12105	AMD33_RS01560
Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	BEN15_RS06060	A9497_RS09055	STER_RS03075	STU_RS12450	AMD33_RS01910
Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	BEN15_RS09825	A9497_RS02760	STER_RS06510	STU_RS15775	AMD33_RS05455
Amino Acid-Polyamine-Organocation (APC)		amino acid specific permease	-	BEN15_RS00750	A9497_RS03620	STER_RS07395	STU_RS16585	AMD33_RS06305
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (glutamate:GABA antiporter)	*gadC*	BEN15_RS05205	-	-	-	-
Amino Acid-Polyamine-Organocation (APC)		histidine:histamine antiporter	*hdcP*	BEN15_RS07915	-	-	-	-
Branched Chain Amino Acid Exporter (LIV-E)		branched-chain amino acid permease (AzlC)	-	BEN15_RS01775	A9497_RS04690	STER_RS08520	STU_RS17670	AMD33_RS07365
Resistance to Homoserine/Threonine (RhtB)		LysE family translocator (amino acid efflux protein)	-	BEN15_RS09935	A9497_RS02865	STER_RS06635	STU_RS15885	AMD33_RS05560
Proton-dependent Oligopeptide Transporter (POT)	di-tripeptide transporter	peptide MFS transporter	*dtpT*	BEN15_RS08025	A9497_RS01085	STER_RS04850	STU_RS14105	AMD33_RS03735
Branched Chain Amino Acid:Cation Symporter (LIVCS)	transporter (symporter)	branched-chain amino acid transport system II carrier protein	*brnQ*	BEN15_RS09805	A9497_RS02740	STER_RS06490	STU_RS15755	AMD33_RS05435
Alanine or Glycine:Cation Symporter (AGCS)	transporter (symporter)	sodium:alanine symporter family protein	-	BEN15_RS08080	A9497_RS01125	STER_RS04890	STU_RS14155	AMD33_RS03680
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		Na+/serine-threonine symporter	*sstT*	BEN15_RS04645	A9497_RS07645	STER_RS01675	STU_RS11125	AMD33_RS00550

Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		dicarboxylate/amino acid:cation symporter	-	BEN15_RS01025	-	-	-	AMD33_RS08020

Transporter family	**product**		**gene**	**MN-BM-A02**	**MN-ZLW-002**	**NCTC12958**	**ND03**	**ND07**
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	Oligopeptide transport ATP-binding protein OppF	*OppF*	MNA02_RS06830	Y1U_RS06850	DQL34_RS07995	STND_RS06810	BBD27_RS07210
		Oligopeptide transport ATP-binding protein OppD	*OppD*	MNA02_RS06835	Y1U_RS06855	DQL34_RS08000	STND_RS06815	BBD27_RS07205
		Oligopeptide transport system permease protein OppC	*OppC*	MNA02_RS06840	Y1U_RS06860	DQL34_RS08005	STND_RS06820	BBD27_RS07200
		Oligopeptide transport system permease protein OppB	*OppB*	MNA02_RS06845	Y1U_RS06865	DQL34_RS08010	STND_RS06825	BBD27_RS07195
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	MNA02_RS06850	Y1U_RS06870	-	STND_RS06830	BBD27_RS07190
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	MNA02_RS06860	Y1U_RS06880	DQL34_RS08015	STND_RS06845	BBD27_RS07180
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	peptide ABC transporter substrate-binding protein	*OppA*	-	-	-	-	-
		ABC transporter permease	*OppB*	-	-	-	-	-
		ABC transporter permease	*OppC*	-	-	-	-	-
		ABC transporter ATP-binding protein	*OppD*	-	-	-	-	-
ATP-binding Cassette (ABC)	branched-chain amino acids ABC transporter	branched-chain amino acid ABC transporter substrate-binding protein	*livK*	MNA02_RS01820	Y1U_RS01845	DQL34_RS02190	STND_RS01880	BBD27_RS02530
		branched-chain amino acid ABC transporter permease	*livH*	MNA02_RS01825	Y1U_RS01850	DQL34_RS02195	STND_RS01885	BBD27_RS02525
		branched-chain amino acid ABC transporter permease	*livM*	MNA02_RS01830	Y1U_RS01855	DQL34_RS02200	STND_RS01890	BBD27_RS02520
		ABC transporter ATP-binding protein	*livG*	MNA02_RS01835	Y1U_RS01860	DQL34_RS02205	STND_RS01895	BBD27_RS02515
		ABC transporter ATP-binding protein	*livF*	MNA02_RS01840	Y1U_RS01865	DQL34_RS02210	STND_RS01900	BBD27_RS02510

ATP-binding Cassette (ABC)	glutamine ABC transporter	amino acid ABC transporter permease	-	MNA02_RS07100	Y1U_RS07115	DQL34_RS08250	STND_RS07105	BBD27_RS06945
		amino acid ABC transporter ATP-binding protein	-	MNA02_RS07105	Y1U_RS07120	DQL34_RS08255	STND_RS07110	BBD27_RS06940
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter permease	-	MNA02_RS04395	Y1U_RS05115	DQL34_RS05475	STND_RS04320	BBD27_RS09630
		amino acid ABC transporter ATP-binding protein	-	MNA02_RS04400	Y1U_RS05110	DQL34_RS05480	STND_RS04325	BBD27_RS09625
		amino acid ABC transporter substrate-binding protein	-	MNA02_RS04405	Y1U_RS05105	DQL34_RS05485	STND_RS04330	BBD27_RS09620
ATP-binding Cassette (ABC)	polar amino acid (cysteine) ABC transporter	amino acid ABC transporter substrate-binding protein	-	MNA02_RS07480	Y1U_RS07495	DQL34_RS08625	STND_RS07485	BBD27_RS06560
		amino acid ABC transporter ATP-binding protein	-	MNA02_RS07485	Y1U_RS07500	DQL34_RS08630	STND_RS07490	BBD27_RS06555
		amino acid ABC transporter permease	-	MNA02_RS07490	Y1U_RS07505	DQL34_RS08635	STND_RS07495	BBD27_RS06550
		amino acid ABC transporter permease	-	MNA02_RS07495	Y1U_RS07510	DQL34_RS08640	STND_RS07500	BBD27_RS06545
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	MNA02_RS07840	Y1U_RS07855	DQL34_RS09085	STND_RS07845	BBD27_RS06200
		amino acid ABC transporter permease	-	MNA02_RS07845	Y1U_RS07860	DQL34_RS09090	STND_RS07850	BBD27_RS06195
		amino acid ABC transporter substrate-binding protein	-	MNA02_RS07850	Y1U_RS07865	DQL34_RS09095	STND_RS07855	BBD27_RS06190
ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	MNA02_RS05460	Y1U_RS04010	DQL34_RS06495	STND_RS05410	BBD27_RS08565
		glutamine ABC transporter substrate-binding protein	-	MNA02_RS05465	Y1U_RS04005	DQL34_RS06500	STND_RS05415	BBD27_RS08560
		amino acid ABC transporter permease	-	MNA02_RS05470	Y1U_RS04000	DQL34_RS06505	STND_RS05420	BBD27_RS08555
		amino acid ABC transporter permease	-	MNA02_RS05475	Y1U_RS03995	DQL34_RS06510	STND_RS05425	BBD27_RS08550
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	-	-	DQL34_RS07505	-	-
		amino acid ABC transporter permease	-	-	-	DQL34_RS07510	-	-
		amino acid ABC transporter permease	-	-	-	DQL34_RS07515	-	-
		amino acid ABC transporter substrate-binding protein	-	-	-	DQL34_RS07520	-	-

ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter substrate-binding protein	-	-	-	DQL34_RS07410	-	-
		amino acid ABC transporter ATP-binding protein	-	-	-	DQL34_RS07420	-	-
		amino acid ABC transporter permease	-	-	-	DQL34_RS07425	-	-
ATP-binding Cassette (ABC)	spermidine/putrescine ABC transporter	spermidine/putrescine ABC transporter, periplasmic binding protein PotD	*potD*	MNA02_RS07260	Y1U_RS07275	DQL34_RS08405	STND_RS07265	BBD27_RS06780
		spermidine/putrescine ABC transporter, permease component potC	*potC*	MNA02_RS07265	Y1U_RS07280	DQL34_RS08410	STND_RS07270	BBD27_RS06775
		spermidine/putrescine ABC transporter, permease component PotB	*potB*	MNA02_RS07270	Y1U_RS07285	DQL34_RS08415	STND_RS07275	BBD27_RS06770
		Putrescine transport ATP-binding protein PotA	*potA*	MNA02_RS07275	Y1U_RS07290	DQL34_RS08420	STND_RS07280	BBD27_RS06765
ATP-binding Cassette (ABC)	methionine ABC transporter	amino acid ABC transporter substrate protein	*metQ*	MNA02_RS01540	Y1U_RS01575	DQL34_RS01895	STND_RS01605	BBD27_RS02805
		methionine import ATP-binding protein MetN	*metN*	MNA02_RS01550	Y1U_RS01585	DQL34_RS01905	STND_RS01615	BBD27_RS02795
		methionine ABC transporter permease	*metP*	MNA02_RS01555	Y1U_RS01590	DQL34_RS01910	STND_RS01620	BBD27_RS02790
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (asparagine)	-	MNA02_RS02545	Y1U_RS02600	DQL34_RS03055	STND_RS02655	BBD27_RS01805
Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	MNA02_RS02955	Y1U_RS02955	DQL34_RS03615	STND_RS03025	BBD27_RS01375
Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	MNA02_RS06440	Y1U_RS06465	DQL34_RS07480	STND_RS06425	BBD27_RS07580
Amino Acid-Polyamine-Organocation (APC)		amino acid specific permease	-	MNA02_RS07310	Y1U_RS07325	DQL34_RS08455	STND_RS07315	BBD27_RS06730
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (glutamate:GABA antiporter)	*gadC*	-	-	-	STND_RS02110	-
Amino Acid-Polyamine-Organocation (APC)		histidine:histamine antiporter	*hdcP*	-	-	-	-	-
Branched Chain Amino Acid Exporter (LIV-E)		branched-chain amino acid permease (AzlC)	-	MNA02_RS08390	Y1U_RS08385	DQL34_RS09865	STND_RS08370	BBD27_RS05655

Resistance to Homoserine/Threonine (RhtB)		LysE family translocator (amino acid efflux protein)	-	MNA02_RS06550	Y1U_RS06570	DQL34_RS04455	STND_RS06530	BBD27_RS07475
Proton-dependent Oligopeptide Transporter (POT)	di-tripeptide transporter	peptide MFS transporter	*dtpT*	MNA02_RS04760	Y1U_RS04750	DQL34_RS05825	STND_RS04690	BBD27_RS09265
Branched Chain Amino Acid:Cation Symporter (LIVCS)	transporter (symporter)	branched-chain amino acid transport system II carrier protein	*brnQ*	MNA02_RS06420	Y1U_RS06445	DQL34_RS07455	STND_RS06405	BBD27_RS07600
Alanine or Glycine:Cation Symporter (AGCS)	transporter (symporter)	sodium:alanine symporter family protein	-	MNA02_RS04800	Y1U_RS04695	DQL34_RS05865	STND_RS04745	BBD27_RS09225
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		Na+/serine-threonine symporter	*sstT*	MNA02_RS01560	Y1U_RS01595	DQL34_RS01915	STND_RS01625	BBD27_RS02785
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		dicarboxylate/amino acid:cation symporter	-	-	Y1U_RS09030	DQL34_RS08770	STND_RS09025	-

Transporter family	**product**		**gene**	**S9**	**SMQ-301**	**ST3**
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	Oligopeptide transport ATP-binding protein OppF	*OppF*	AVT04_RS07760	SMQ301_RS06895	BGL51_RS07110
		Oligopeptide transport ATP-binding protein OppD	*OppD*	AVT04_RS07755	SMQ301_RS06900	BGL51_RS07115
		Oligopeptide transport system permease protein OppC	*OppC*	AVT04_RS07750	SMQ301_RS06905	BGL51_RS07120
		Oligopeptide transport system permease protein OppB	*OppB*	AVT04_RS07745	SMQ301_RS06910	BGL51_RS07125
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	AVT04_RS07740	SMQ301_RS06915	BGL51_RS07130
		Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA	*OppA*	AVT04_RS07730	SMQ301_RS06925	BGL51_RS07140
ATP-binding Cassette (ABC)	oligopeptide opp ABC transporter	peptide ABC transporter substrate-binding protein	*OppA*	-	SMQ301_RS00690	BGL51_RS00655
		ABC transporter permease	*OppB*	-	SMQ301_RS00695	BGL51_RS00660

		ABC transporter permease	*OppC*	-	SMQ301_RS00700	BGL51_RS00665
		ABC transporter ATP-binding protein	*OppD*	-	SMQ301_RS00705	BGL51_RS00670
ATP-binding Cassette (ABC)	branched-chain amino acids ABC transporter	branched-chain amino acid ABC transporter substrate-binding protein	*livK*	AVT04_RS03100	SMQ301_RS01955	BGL51_RS01915
		branched-chain amino acid ABC transporter permease	*livH*	AVT04_RS03095	SMQ301_RS01960	BGL51_RS01920
		branched-chain amino acid ABC transporter permease	*livM*	AVT04_RS03090	SMQ301_RS01965	BGL51_RS01925
		ABC transporter ATP-binding protein	*livG*	AVT04_RS03085	SMQ301_RS01970	BGL51_RS01930
		ABC transporter ATP-binding protein	*livF*	AVT04_RS03080	SMQ301_RS01975	BGL51_RS01935
ATP-binding Cassette (ABC)	glutamine ABC transporter	amino acid ABC transporter permease	-	AVT04_RS07500	SMQ301_RS07175	BGL51_RS07390
		amino acid ABC transporter ATP-binding protein	-	AVT04_RS07495	SMQ301_RS07180	BGL51_RS07395
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter permease	-	AVT04_RS00705	SMQ301_RS04460	BGL51_RS04545
		amino acid ABC transporter ATP-binding protein	-	AVT04_RS00700	SMQ301_RS04465	BGL51_RS04550
		amino acid ABC transporter substrate-binding protein	-	AVT04_RS00695	SMQ301_RS04470	BGL51_RS04555
ATP-binding Cassette (ABC)	polar amino acid (cysteine) ABC transporter	amino acid ABC transporter substrate-binding protein	-	AVT04_RS07120	SMQ301_RS07550	BGL51_RS07775
		amino acid ABC transporter ATP-binding protein	-	AVT04_RS07115	SMQ301_RS07555	BGL51_RS07780
		amino acid ABC transporter permease	-	AVT04_RS07110	SMQ301_RS07560	BGL51_RS07785
		amino acid ABC transporter permease	-	AVT04_RS07105	SMQ301_RS07565	BGL51_RS07790
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	AVT04_RS06760	SMQ301_RS07920	BGL51_RS08155
		amino acid ABC transporter permease	-	AVT04_RS06755	SMQ301_RS07925	BGL51_RS08160
		amino acid ABC transporter substrate-binding protein	-	AVT04_RS06750	SMQ301_RS07930	BGL51_RS08165
ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	AVT04_RS09045	SMQ301_RS05520	BGL51_RS05745

		glutamine ABC transporter substrate-binding protein	-	AVT04_RS09040	SMQ301_RS05525	BGL51_RS05750
		amino acid ABC transporter permease	-	AVT04_RS09035	SMQ301_RS05530	BGL51_RS05755
		amino acid ABC transporter permease	-	AVT04_RS09030	SMQ301_RS05535	BGL51_RS05760
ATP-binding Cassette (ABC)	polar amino acid ABC transporter	amino acid ABC transporter ATP-binding protein	-	-	-	-
		amino acid ABC transporter permease	-	-	-	-
		amino acid ABC transporter permease	-	-	-	-
		amino acid ABC transporter substrate-binding protein	-	-	-	-
ATP-binding Cassette (ABC)	amino acid ABC transporter	amino acid ABC transporter substrate-binding protein	-	AVT04_RS08160	-	-
		amino acid ABC transporter ATP-binding protein	-	AVT04_RS08150	-	-
		amino acid ABC transporter permease	-	AVT04_RS08145	-	-
ATP-binding Cassette (ABC)	spermidine/putrescine ABC transporter	spermidine/putrescine ABC transporter, periplasmic binding protein PotD	*potD*	AVT04_RS07340	SMQ301_RS07330	BGL51_RS07545
		spermidine/putrescine ABC transporter, permease component potC	*potC*	AVT04_RS07335	SMQ301_RS07335	BGL51_RS07550
		spermidine/putrescine ABC transporter, permease component PotB	*potB*	AVT04_RS07330	SMQ301_RS07340	BGL51_RS07555
		Putrescine transport ATP-binding protein PotA	*potA*	AVT04_RS07325	SMQ301_RS07345	BGL51_RS07560
ATP-binding Cassette (ABC)	methionine ABC transporter	amino acid ABC transporter substrate protein	*metQ*	AVT04_RS03370	SMQ301_RS01670	BGL51_RS01635
		methionine import ATP-binding protein MetN	*metN*	AVT04_RS03360	SMQ301_RS01680	BGL51_RS01645
		methionine ABC transporter permease	*metP*	AVT04_RS03355	SMQ301_RS01685	BGL51_RS01650
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (asparagine)	-	AVT04_RS02365	SMQ301_RS02670	BGL51_RS02625

Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	AVT04_RS02015	SMQ301_RS03060	BGL51_RS03050
Amino Acid-Polyamine-Organocation (APC)		amino acid permease	-	AVT04_RS08085	SMQ301_RS06520	BGL51_RS06720
Amino Acid-Polyamine-Organocation (APC)		amino acid specific permease	-	AVT04_RS07290	SMQ301_RS07380	BGL51_RS07595
Amino Acid-Polyamine-Organocation (APC)		amino acid permease (glutamate:GABA antiporter)	*gadC*	-	-	-
Amino Acid-Polyamine-Organocation (APC)		histidine:histamine antiporter	*hdcP*	-	-	-
Branched Chain Amino Acid Exporter (LIV-E)		branched-chain amino acid permease (AzlC)	-	AVT04_RS06230	SMQ301_RS08510	BGL51_RS08620
Resistance to Homoserine/Threonine (RhtB)		LysE family translocator (amino acid efflux protein)	-	AVT04_RS07985	SMQ301_RS06620	BGL51_RS06845
Proton-dependent Oligopeptide Transporter (POT)	di-tripeptide transporter	peptide MFS transporter	*dtpT*	AVT04_RS00295	SMQ301_RS04830	BGL51_RS05050
Branched Chain Amino Acid:Cation Symporter (LIVCS)	transporter (symporter)	branched-chain amino acid transport system II carrier protein	*brnQ*	AVT04_RS08105	SMQ301_RS06500	BGL51_RS06700
Alanine or Glycine:Cation Symporter (AGCS)	transporter (symporter)	sodium:alanine symporter family protein	-	AVT04_RS00245	SMQ301_RS04870	BGL51_RS05090
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		Na+/serine-threonine symporter	*sstT*	AVT04_RS03350	SMQ301_RS01690	BGL51_RS01655
Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS)		dicarboxylate/amino acid:cation symporter	-	-	-	BGL51_RS09235

Supplementary Table S6. Genes involved in amino acid biosynthesis identified in the 23 *S. thermophilus* strains according to Kyoto Encyclopedia of Genes and Genomes (KEGG) database and Hols et al. (2005). Genes highlighted in red correspond to putative pseudogenes

amino acid	product	gene	ACA-DC 2	APC151	ASCC 1275	B59671	CNRZ1066
Serine biosynthesis	3-phosphoglycerate dehydrogenase	*serA*	STACADC2_RS06855	B1761_RS03130	T303_RS08320	CG712_RS03650	STR_RS07155
	3-phosphoserine/phosphohydroxythreonine transaminase	*serC*	STACADC2_RS06865	B1761_RS03140	T303_RS08330	CG712_RS03660	STR_RS07165
	phosphoserine phosphatase SerB	*serB*	STACADC2_RS06810	B1761_RS03085	T303_RS08275	CG712_RS03605	STR_RS07110
Threonine biosynthesis	aspartate kinase	*lysC*	STACADC2_RS01955	B1761_RS07670	T303_RS03040	CG712_RS07990	STR_RS02025
	aspartate-semialdehyde dehydrogenase	*asd*	STACADC2_RS05960	B1761_RS02100	T303_RS07315	CG712_RS02700	STR_RS06140
	homoserine dehydrogenase	*hom*	STACADC2_RS02370	B1761_RS08125	T303_RS03430	CG712_RS08415	STR_RS02435
	homoserine kinase	*thrB*	STACADC2_RS02375	B1761_RS08130	T303_RS03435	CG712_RS08420	STR_RS02440
	threonine synthase	*thrC*	STACADC2_RS08540	B1761_RS04840	T303_RS00330	CG712_RS05340	STR_RS08820
Cysteine biosynthesis	serine O-acetyltransferase	*cysE*	STACADC2_RS00550	B1761_RS06175	T303_RS01660	CG712_RS06570	STR_RS00570
	serine acetyltransferase	*cysE*	STACADC2_RS04085	B1761_RS00035	T303_RS10265	-	STR_RS04245
	serine acetyltransferase	*cysE*	-	-	-	-	-
	cysteine synthase A	*cysK*	STACADC2_RS01850	B1761_RS07565	T303_RS02945	CG712_RS07885	STR_RS01915
	cysteine synthase family protein	*cysK*	STACADC2_RS04075	B1761_RS00025	T303_RS05385	-	STR_RS04235
Methionine biosynthesis	aspartate kinase	*lysC*	STACADC2_RS01955	B1761_RS07670	T303_RS03040	CG712_RS07990	STR_RS02025
	aspartate-semialdehyde dehydrogenase	*asd*	STACADC2_RS05960	B1761_RS02100	T303_RS07315	CG712_RS02700	STR_RS06140
	homoserine dehydrogenase	*hom*	STACADC2_RS02370	B1761_RS08125	T303_RS03430	CG712_RS08415	STR_RS02435
	homoserine O-succinyltransferase	*metA*	STACADC2_RS05590	B1761_RS01670	T303_RS06865	CG712_RS02330	STR_RS05785
	cystathionine gamma-synthase	*metB*	STACADC2_RS01780	B1761_RS07495	T303_RS02870	CG712_RS07815	STR_RS01845

	aminotransferase class V-fold PLP-dependent enzyme	*metC*	STACADC2_RS04080	B1761_RS00030	T303_RS05390	-	STR_RS04240
	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	*metE*	STACADC2_RS03815	B1761_RS09670	T303_RS05075	CG712_RS00430	STR_RS03980
Histidine biosynthesis	ATP phosphoribosyltransferase	*hisG*	-	B1761_RS01745	T303_RS06945	-	-
	ATP phosphoribosyltransferase regulatory subunit	*hisZ*	-	B1761_RS01750	T303_RS06950	-	-
	phosphoribosyl-ATP diphosphatase	*hisE*	-	B1761_RS01710	T303_RS06910	-	-
	phosphoribosyl-AMP cyclohydrolase	*hisI*	-	B1761_RS01715	T303_RS06915	-	-
	1-(5-phosphoribosyl)-5-((5-phosphoribosylamino)methylideneamino)imidazole-4- carboxamide isomerase	*hisA*	-	B1761_RS01725	T303_RS06925	-	-
	imidazole glycerol phosphate synthase subunit HisH	*hisH*	-	B1761_RS01730	T303_RS06930	-	-
	imidazole glycerol phosphate synthase subunit HisF	*hisF*	-	B1761_RS01720	T303_RS06920	-	-
	imidazoleglycerol-phosphate dehydratase	*hisB*	-	B1761_RS01735	T303_RS06935	-	-
	histidinol-phosphate transaminase	*hisC*	-	B1761_RS01755	T303_RS06955	-	-
	hypothetical protein	*hisK*	STACADC2_RS05640	B1761_RS01780	T303_RS06985	CG712_RS02375	STR_RS05835
	histidinol dehydrogenase	*hisD*	-	B1761_RS01740	T303_RS06940	-	-

Glycine biosynthesis	serine hydroxymethyltransferase	*glyA*	STACADC2_RS03745	B1761_RS09510	T303_RS04920	CG712_RS00360	STR_RS03825
Proline biosynthesis	glutamate-5-semialdehyde dehydrogenase	*proA*	STACADC2_RS07720	B1761_RS04010	T303_RS09200	CG712_RS04520	STR_RS08010
	glutamate 5-kinase	*proB*	STACADC2_RS07725	B1761_RS04015	T303_RS09205	CG712_RS04525	STR_RS08015
	pyrroline-5-carboxylate reductase	*proC*	STACADC2_RS08410	B1761_RS04705	T303_RS00195	CG712_RS05210	STR_RS08690
Glutamine biosynthesis	type I glutamate--ammonia ligase	*glnA*	STACADC2_RS08040	B1761_RS04325	T303_RS09515	CG712_RS04840	STR_RS08325
Arginine biosynthesis	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	STACADC2_RS02350	B1761_RS08105	T303_RS03410	CG712_RS08395	STR_RS02415
	acetylglutamate kinase	*argB*	STACADC2_RS02355	B1761_RS08110	T303_RS03415	CG712_RS08400	STR_RS02420
	N-acetyl-gamma-glutamyl-phosphate reductase	*argC*	STACADC2_RS02345	B1761_RS08100	T303_RS03405	CG712_RS08390	STR_RS02410
	acetylornithine transaminase	*argD*	STACADC2_RS02360	B1761_RS08115	T303_RS03420	CG712_RS08405	STR_RS02425
	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	STACADC2_RS02350	B1761_RS08105	T303_RS03410	CG712_RS08395	STR_RS02415
	ornithine carbamoyltransferase	*argF*	STACADC2_RS03015	B1761_RS08790	T303_RS04140	CG712_RS09065	STR_RS03095
	argininosuccinate synthase	*argG*	STACADC2_RS08230	B1761_RS04520	T303_RS00025	CG712_RS05030	STR_RS08520
	argininosuccinate lyase	*argH*	STACADC2_RS08225	B1761_RS04515	T303_RS00020	CG712_RS05025	STR_RS08515
Asparagine biosynthesis (from aspartate)	asparagine synthetase A	*asnA*	STACADC2_RS01950	B1761_RS07665	T303_RS03035	CG712_RS07985	STR_RS02020
Valine biosynthesis	acetolactate synthase AlsS	*ilvB*	STACADC2_RS04400	B1761_RS00315	T303_RS05670	CG712_RS00960	STR_RS04555
	acetolactate synthase I large subunit	*ilvB*	STACADC2_RS08515	B1761_RS04815	T303_RS00305	CG712_RS05315	STR_RS08795
	acetolactate synthase small subunit	*ilvN*	STACADC2_RS08510	B1761_RS04810	T303_RS00300	CG712_RS05310	STR_RS08790

	ketol-acid reductoisomerase	*ilvC*	STACADC2_RS08505	B1761_RS04805	T303_RS00295	CG712_RS05305	STR_RS08785
	dihydroxy-acid dehydratase	*ilvD*	STACADC2_RS07240	B1761_RS03510	T303_RS08700	CG712_RS04030	STR_RS07535
	dihydroxy-acid dehydratase	*ilvD*	STACADC2_RS08520	B1761_RS04820	T303_RS00310	CG712_RS05320	STR_RS08800
	branched-chain amino acid aminotransferase	*ilvE*	STACADC2_RS02935	B1761_RS08715	T303_RS04070	CG712_RS08995	STR_RS03010
Isoleucine biosynthesis	threonine dehydratase	*ilvA*	STACADC2_RS00865	B1761_RS06585	T303_RS01975	CG712_RS06955	STR_RS00970
	acetolactate synthase AlsS	*ilvB*	STACADC2_RS04400	B1761_RS00315	T303_RS05670	CG712_RS00960	STR_RS04555
	acetolactate synthase I large subunit	*ilvB*	STACADC2_RS08515	B1761_RS04815	T303_RS00305	CG712_RS05315	STR_RS08795
	acetolactate synthase small subunit	*ilvN*	STACADC2_RS08510	B1761_RS04810	T303_RS00300	CG712_RS05310	STR_RS08790
	ketol-acid reductoisomerase	*ilvC*	STACADC2_RS08505	B1761_RS04805	T303_RS00295	CG712_RS05305	STR_RS08785
	dihydroxy-acid dehydratase	*ilvD*	STACADC2_RS07240	B1761_RS03510	T303_RS08700	CG712_RS04030	STR_RS07535
	dihydroxy-acid dehydratase	*ilvD*	STACADC2_RS08520	B1761_RS04820	T303_RS00310	CG712_RS05320	STR_RS08800
	branched-chain amino acid aminotransferase	*ilvE*	STACADC2_RS02935	B1761_RS08715	T303_RS04070	CG712_RS08995	STR_RS03010
Leucine biosynthesis	acetolactate synthase AlsS	*ilvB*	STACADC2_RS04400	B1761_RS00315	T303_RS05670	CG712_RS00960	STR_RS04555
	acetolactate synthase I large subunit	*ilvB*	STACADC2_RS08515	B1761_RS04815	T303_RS00305	CG712_RS05315	STR_RS08795
	acetolactate synthase small subunit	*ilvN*	STACADC2_RS08510	B1761_RS04810	T303_RS00300	CG712_RS05310	STR_RS08790
	ketol-acid reductoisomerase	*ilvC*	STACADC2_RS08505	B1761_RS04805	T303_RS00295	CG712_RS05305	STR_RS08785
	dihydroxy-acid dehydratase	*ilvD*	STACADC2_RS07240	B1761_RS03510	T303_RS08700	CG712_RS04030	STR_RS07535
	2-isopropylmalate synthase	*leuA*	STACADC2_RS05505	B1761_RS01585	T303_RS06780	CG712_RS02245	STR_RS05690
	3-isopropylmalate dehydratase large subunit	*leuC*	STACADC2_RS05490	B1761_RS01570	T303_RS06765	CG712_RS02230	STR_RS05675

	3-isopropylmalate dehydratase small subunit	*leuD*	STACADC2_RS05485	B1761_RS01565	T303_RS06760	CG712_RS02225	STR_RS05670
	3-isopropylmalate dehydrogenase	*leuB*	STACADC2_RS05500	B1761_RS01580	T303_RS06775	CG712_RS02240	STR_RS05685
	branched-chain amino acid aminotransferase	*ilvE*	STACADC2_RS02935	B1761_RS08715	T303_RS04070	CG712_RS08995	STR_RS03010
Tryptophan biosynthesis	3-deoxy-7-phosphoheptulonate synthase	*aroG1*	STACADC2_RS07810	B1761_RS04095	T303_RS09285	CG712_RS04610	-
	3-deoxy-7-phosphoheptulonate synthase	*aroG2*	STACADC2_RS07815	B1761_RS04100	T303_RS09290	CG712_RS04615	STR_RS08095
	3-dehydroquinate synthase	*aroB*	STACADC2_RS03230	B1761_RS09000	T303_RS04335	CG712_RS09270	STR_RS03275
	type I 3-dehydroquinate dehydratase	*aroD*	STACADC2_RS03220	B1761_RS08990	T303_RS04325	CG712_RS09260	STR_RS03265
	shikimate dehydrogenase	*aroE*	STACADC2_RS03225	B1761_RS08995	T303_RS04330	CG712_RS09265	STR_RS03270
	shikimate kinase	*aroK*	STACADC2_RS03260	B1761_RS09030	T303_RS04365	CG712_RS09300	STR_RS03305
	3-phosphoshikimate 1-carboxyvinyltransferase	*aroA*	STACADC2_RS03255	B1761_RS09025	T303_RS04360	CG712_RS09295	STR_RS03300
	chorismate synthase	*aroC*	STACADC2_RS03235	B1761_RS09005	T303_RS04340	CG712_RS09275	STR_RS03280
	aminodeoxychorismate/anthranilate synthase component II	*trpG*	STACADC2_RS07180	B1761_RS03450	T303_RS08640	CG712_RS03970	STR_RS07475
	anthranilate synthase component I	*trpE*	STACADC2_RS07185	B1761_RS03455	T303_RS08645	CG712_RS03975	STR_RS07480
	anthranilate phosphoribosyltransferase	*trpD*	STACADC2_RS07175	B1761_RS03445	T303_RS08635	CG712_RS03965	STR_RS07470
	phosphoribosylanthranilate isomerase	*trpF*	STACADC2_RS07165	B1761_RS03435	T303_RS08625	CG712_RS03955	STR_RS07460
	indole-3-glycerol phosphate synthase TrpC	*trpC*	STACADC2_RS07170	B1761_RS03440	T303_RS08630	CG712_RS03960	STR_RS07465
	tryptophan synthase subunit alpha	*trpA*	STACADC2_RS07155	B1761_RS03425	T303_RS08615	CG712_RS03945	STR_RS07450

	tryptophan synthase subunit beta	*trpB*	STACADC2_RS07160	B1761_RS03430	T303_RS08620	CG712_RS03950	STR_RS07455
Phenylalanine biosynthesis	chorismate mutase	*aroH*	STACADC2_RS05390	B1761_RS01460	T303_RS06655	CG712_RS02130	STR_RS05570
	prephenate dehydratase	*pheA2*	STACADC2_RS03265	B1761_RS09035	T303_RS04370	CG712_RS09305	STR_RS03310
	pyridoxal phosphate-dependent aminotransferase		STACADC2_RS03945	B1761_RS09800	T303_RS05255	CG712_RS00560	STR_RS04115
Tyrosine biosynthesis	chorismate mutase	*aroH*	STACADC2_RS05390	B1761_RS01460	T303_RS06655	CG712_RS02130	STR_RS05570
	prephenate dehydrogenase	*tyrA*	STACADC2_RS03240	B1761_RS09010	T303_RS04345	CG712_RS09280	STR_RS03285
	pyridoxal phosphate-dependent aminotransferase		STACADC2_RS03945	B1761_RS09800	T303_RS05255	CG712_RS00560	STR_RS04115
Alanine biosynthesis	pyridoxal phosphate-dependent aminotransferase		STACADC2_RS07400	B1761_RS03670	T303_RS08855	CG712_RS04190	STR_RS07690
Aspartate biosynthesis	pyridoxal phosphate-dependent aminotransferase		STACADC2_RS03945	B1761_RS09800	T303_RS05255	CG712_RS00560	STR_RS04115
Glutamate biosynthesis	pyridoxal phosphate-dependent aminotransferase		STACADC2_RS03945	B1761_RS09800	T303_RS05255	CG712_RS00560	STR_RS04115
	pyridoxal phosphate-dependent aminotransferase		STACADC2_RS07400	B1761_RS03670	T303_RS08855	CG712_RS04190	STR_RS07690
Lysine biosynthesis (succinyl-Dap pathway)	aspartate kinase	*lysC*	STACADC2_RS01955	B1761_RS07670	T303_RS03040	CG712_RS07990	STR_RS02025
	aspartate-semialdehyde dehydrogenase	*asd*	STACADC2_RS05960	B1761_RS02100	T303_RS07315	CG712_RS02700	STR_RS06140
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	STACADC2_RS05955	B1761_RS02095	T303_RS07310	CG712_RS02695	STR_RS06135
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	STACADC2_RS02165	B1761_RS07915	T303_RS03220	CG712_RS08200	STR_RS02215
	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase	*dapD*	STACADC2_RS08335	B1761_RS04630	T303_RS00130	CG712_RS05140	STR_RS08625

	acetylornithine transaminase	*argD*	STACADC2_RS02360	B1761_RS08115	T303_RS03420	CG712_RS08405	STR_RS02425
	succinyl-diaminopimelate desuccinylase	*dapE*	-	-	-	-	-
	diaminopimelate epimerase	*dapF*	-	-	-	-	-
	diaminopimelate decarboxylase	*lysA*	STACADC2_RS01325	B1761_RS07040	T303_RS02420	CG712_RS07395	STR_RS01425
Lysine biosynthesis (acetyl-Dap pathway)	aspartate kinase	*lysC*	STACADC2_RS01955	B1761_RS07670	T303_RS03040	CG712_RS07990	STR_RS02025
	aspartate-semialdehyde dehydrogenase	*asd*	STACADC2_RS05960	B1761_RS02100	T303_RS07315	CG712_RS02700	STR_RS06140
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	STACADC2_RS05955	B1761_RS02095	T303_RS07310	CG712_RS02695	STR_RS06135
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	STACADC2_RS02165	B1761_RS07915	T303_RS03220	CG712_RS08200	STR_RS02215
	tetrahydrodipicolinate N-acetyltransferase	*dapH*	-	-	-	-	-
	pyridoxal phosphate-dependent aminotransferase	*patA*	STACADC2_RS00210	B1761_RS05805	T303_RS01310	CG712_RS06230	STR_RS00225
	N-acetyldiaminopimelate deacetylase	*dapL*	STACADC2_RS08330	B1761_RS04625	T303_RS00125	CG712_RS05135	STR_RS08620
	diaminopimelate epimerase	*dapF*	-	-	-	-	-
	diaminopimelate decarboxylase	*lysA*	STACADC2_RS01325	B1761_RS07040	T303_RS02420	CG712_RS07395	STR_RS01425

amino acid	**product**	**gene**	**CS8**	**DGCC 7710**	**EPS**	**GABA**	**JIM 8232**
Serine biosynthesis	3-phosphoglycerate dehydrogenase	*serA*	BAY21_RS07210	CW339_RS07330	CR922_RS02210	CR921_RS07410	STH8232_RS07545
	3-phosphoserine/phosphohydroxythreonine transaminase	*serC*	BAY21_RS07200	CW339_RS07340	CR922_RS02200	CR921_RS07420	STH8232_RS07555

	phosphoserine phosphatase SerB	*serB*	BAY21_RS07255	CW339_RS07285	CR922_RS02255	CR921_RS07365	STH8232_RS07500
Threonine biosynthesis	aspartate kinase	*lysC*	BAY21_RS02875	CW339_RS01930	CR922_RS07425	CR921_RS01995	STH8232_RS02160
	aspartate-semialdehyde dehydrogenase	*asd*	BAY21_RS08220	CW339_RS06305	CR922_RS03265	CR921_RS06300	STH8232_RS06600
	homoserine dehydrogenase	*hom*	BAY21_RS02470	CW339_RS02330	CR922_RS07015	CR921_RS02440	STH8232_RS02555
	homoserine kinase	*thrB*	BAY21_RS02465	CW339_RS02335	CR922_RS07010	CR921_RS02445	STH8232_RS02560
	threonine synthase	*thrC*	BAY21_RS05535	CW339_RS09045	CR922_RS08980	CR921_RS09050	STH8232_RS09295
Cysteine biosynthesis	serine O-acetyltransferase	*cysE*	BAY21_RS04275	CW339_RS00535	CR922_RS00540	CR921_RS00545	STH8232_RS00595
	serine acetyltransferase	*cysE*	BAY21_RS00690	CW339_RS04325	CR922_RS05195	CR921_RS04270	STH8232_RS04580
	serine acetyltransferase	*cysE*	-	-	-	-	-
	cysteine synthase A	*cysK*	BAY21_RS02985	CW339_RS01825	CR922_RS07535	CR921_RS01890	STH8232_RS02055
	cysteine synthase family protein	*cysK*	BAY21_RS00700	CW339_RS04315	CR922_RS05205	CR921_RS04260	STH8232_RS04570
Methionine biosynthesis	aspartate kinase	*lysC*	BAY21_RS02875	CW339_RS01930	CR922_RS07425	CR921_RS01995	STH8232_RS02160
	aspartate-semialdehyde dehydrogenase	*asd*	BAY21_RS08220	CW339_RS06305	CR922_RS03265	CR921_RS06300	STH8232_RS06600
	homoserine dehydrogenase	*hom*	BAY21_RS02470	CW339_RS02330	CR922_RS07015	CR921_RS02440	STH8232_RS02555
	homoserine O-succinyltransferase	*metA*	BAY21_RS08570	CW339_RS05865	CR922_RS03625	CR921_RS05875	STH8232_RS06160
	cystathionine gamma-synthase	*metB*	BAY21_RS03055	CW339_RS01755	CR922_RS07605	CR921_RS01820	STH8232_RS01980
	aminotransferase class V-fold PLP-dependent enzyme	*metC*	BAY21_RS00695	CW339_RS04320	CR922_RS05200	CR921_RS04265	STH8232_RS04575
	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	*metE*	BAY21_RS00950	CW339_RS04015	CR922_RS05470	CR921_RS03960	STH8232_RS04275
Histidine biosynthesis	ATP phosphoribosyltransferase	*hisG*	-	CW339_RS05945	-	CR921_RS05955	STH8232_RS06240

	ATP phosphoribosyltransferase regulatory subunit	*hisZ*	-	CW339_RS05950	-	CR921_RS05960	STH8232_RS06245
	phosphoribosyl-ATP diphosphatase	*hisE*	-	CW339_RS05910	-	CR921_RS05920	STH8232_RS06205
	phosphoribosyl-AMP cyclohydrolase	*hisI*	-	CW339_RS05915	-	CR921_RS05925	STH8232_RS06210
	1-(5-phosphoribosyl)-5-((5-phosphoribosylamino)methylideneamino)imidazole-4- carboxamide isomerase	*hisA*	-	CW339_RS05925	-	CR921_RS05935	STH8232_RS06220
	imidazole glycerol phosphate synthase subunit HisH	*hisH*	-	CW339_RS05930	-	CR921_RS05940	STH8232_RS06225
	imidazole glycerol phosphate synthase subunit HisF	*hisF*	-	CW339_RS05920	-	CR921_RS05930	STH8232_RS06215
	imidazoleglycerol-phosphate dehydratase	*hisB*	-	CW339_RS05935	-	CR921_RS05945	STH8232_RS06230
	histidinol-phosphate transaminase	*hisC*	-	CW339_RS05955	-	CR921_RS05965	STH8232_RS06250
	hypothetical protein	*hisK*	BAY21_RS08520	CW339_RS05975	CR922_RS03575	CR921_RS05985	STH8232_RS06285
	histidinol dehydrogenase	*hisD*	-	CW339_RS05940	-	CR921_RS05950	STH8232_RS06235
Glycine biosynthesis	serine hydroxymethyltransferase	*glyA*	BAY21_RS01095	CW339_RS03880	CR922_RS05605	CR921_RS03885	STH8232_RS04170
Proline biosynthesis	glutamate-5-semialdehyde dehydrogenase	*proA*	BAY21_RS06355	CW339_RS08220	CR922_RS01320	CR921_RS08235	STH8232_RS08465
	glutamate 5-kinase	*proB*	BAY21_RS06350	CW339_RS08225	CR922_RS01315	CR921_RS08240	STH8232_RS08470
	pyrroline-5-carboxylate reductase	*proC*	BAY21_RS05665	CW339_RS08910	CR922_RS08850	CR921_RS08915	STH8232_RS09160
Glutamine biosynthesis	type I glutamate--ammonia ligase	*glnA*	BAY21_RS06040	CW339_RS08535	CR922_RS01005	CR921_RS08550	STH8232_RS08780

Arginine biosynthesis	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	BAY21_RS02490	CW339_RS02310	CR922_RS07035	CR921_RS02420	STH8232_RS02535
	acetylglutamate kinase	*argB*	BAY21_RS02485	CW339_RS02315	CR922_RS07030	CR921_RS02425	STH8232_RS02540
	N-acetyl-gamma-glutamyl-phosphate reductase	*argC*	BAY21_RS02495	CW339_RS02305	CR922_RS07040	CR921_RS02415	STH8232_RS02530
	acetylornithine transaminase	*argD*	BAY21_RS02480	CW339_RS02320	CR922_RS07025	CR921_RS02430	STH8232_RS02545
	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	BAY21_RS02490	CW339_RS02310	CR922_RS07035	CR921_RS02420	STH8232_RS02535
	ornithine carbamoyltransferase	*argF*	BAY21_RS01815	CW339_RS03060	CR922_RS06375	CR921_RS03180	STH8232_RS03475
	argininosuccinate synthase	*argG*	BAY21_RS05845	CW339_RS08735	CR922_RS08680	CR921_RS08745	STH8232_RS08990
	argininosuccinate lyase	*argH*	BAY21_RS05850	CW339_RS08730	CR922_RS08675	CR921_RS08740	STH8232_RS08985
Asparagine biosynthesis (from aspartate)	asparagine synthetase A	*asnA*	BAY21_RS02880	CW339_RS01925	CR922_RS07430	CR921_RS01990	STH8232_RS02155
Valine biosynthesis	acetolactate synthase AlsS	*ilvB*	BAY21_RS00375	CW339_RS04615	CR922_RS04870	CR921_RS04560	STH8232_RS04910
	acetolactate synthase I large subunit	*ilvB*	BAY21_RS05560	CW339_RS09020	CR922_RS08955	CR921_RS09025	STH8232_RS09270
	acetolactate synthase small subunit	*ilvN*	BAY21_RS05565	CW339_RS09015	CR922_RS08950	CR921_RS09020	STH8232_RS09265
	ketol-acid reductoisomerase	*ilvC*	BAY21_RS05570	CW339_RS09010	CR922_RS08945	CR921_RS09015	STH8232_RS09260
	dihydroxy-acid dehydratase	*ilvD*	BAY21_RS06835	CW339_RS07705	CR922_RS01830	CR921_RS07790	STH8232_RS07965
	dihydroxy-acid dehydratase	*ilvD*	BAY21_RS05555	CW339_RS09025	CR922_RS08960	CR921_RS09030	STH8232_RS09275
	branched-chain amino acid aminotransferase	*ilvE*	BAY21_RS01900	CW339_RS02990	CR922_RS06445	CR921_RS03110	STH8232_RS03405
Isoleucine biosynthesis	threonine dehydratase	*ilvA*	BAY21_RS03940	CW339_RS00860	CR922_RS08490	CR921_RS00930	STH8232_RS01100
	acetolactate synthase AlsS	*ilvB*	BAY21_RS00375	CW339_RS04615	CR922_RS04870	CR921_RS04560	STH8232_RS04910

	acetolactate synthase I large subunit	*ilvB*	BAY21_RS05560	CW339_RS09020	CR922_RS08955	CR921_RS09025	STH8232_RS09270
	acetolactate synthase small subunit	*ilvN*	BAY21_RS05565	CW339_RS09015	CR922_RS08950	CR921_RS09020	STH8232_RS09265
	ketol-acid reductoisomerase	*ilvC*	BAY21_RS05570	CW339_RS09010	CR922_RS08945	CR921_RS09015	STH8232_RS09260
	dihydroxy-acid dehydratase	*ilvD*	BAY21_RS06835	CW339_RS07705	CR922_RS01830	CR921_RS07790	STH8232_RS07965
	dihydroxy-acid dehydratase	*ilvD*	BAY21_RS05555	CW339_RS09025	CR922_RS08960	CR921_RS09030	STH8232_RS09275
	branched-chain amino acid aminotransferase	*ilvE*	BAY21_RS01900	CW339_RS02990	CR922_RS06445	CR921_RS03110	STH8232_RS03405
Leucine biosynthesis	acetolactate synthase AlsS	*ilvB*	BAY21_RS00375	CW339_RS04615	CR922_RS04870	CR921_RS04560	STH8232_RS04910
	acetolactate synthase I large subunit	*ilvB*	BAY21_RS05560	CW339_RS09020	CR922_RS08955	CR921_RS09025	STH8232_RS09270
	acetolactate synthase small subunit	*ilvN*	BAY21_RS05565	CW339_RS09015	CR922_RS08950	CR921_RS09020	STH8232_RS09265
	ketol-acid reductoisomerase	*ilvC*	BAY21_RS05570	CW339_RS09010	CR922_RS08945	CR921_RS09015	STH8232_RS09260
	dihydroxy-acid dehydratase	*ilvD*	BAY21_RS06835	CW339_RS07705	CR922_RS01830	CR921_RS07790	STH8232_RS07965
	2-isopropylmalate synthase	*leuA*	BAY21_RS08665	CW339_RS05765	CR922_RS03730	CR921_RS05775	STH8232_RS06075
	3-isopropylmalate dehydratase large subunit	*leuC*	BAY21_RS08680	CW339_RS05765	CR922_RS03730	CR921_RS05775	STH8232_RS06060
	3-isopropylmalate dehydratase small subunit	*leuD*	BAY21_RS08685	CW339_RS05760	CR922_RS03735	CR921_RS05770	STH8232_RS06055
	3-isopropylmalate dehydrogenase	*leuB*	BAY21_RS08670	CW339_RS05775	CR922_RS03720	CR921_RS05785	STH8232_RS06070
	branched-chain amino acid aminotransferase	*ilvE*	BAY21_RS01900	CW339_RS02990	CR922_RS06445	CR921_RS03110	STH8232_RS03405
Tryptophan biosynthesis	3-deoxy-7-phosphoheptulonate synthase	*aroG1*	-	CW339_RS08305	CR922_RS01235	CR921_RS08320	STH8232_RS08550

	3-deoxy-7-phosphoheptulonate synthase	*aroG2*	BAY21_RS06270	CW339_RS08310	CR922_RS01230	CR921_RS08325	STH8232_RS08555
	3-dehydroquinate synthase	*aroB*	BAY21_RS01640	CW339_RS03275	CR922_RS06175	CR921_RS03380	STH8232_RS03660
	type I 3-dehydroquinate dehydratase	*aroD*	BAY21_RS01650	CW339_RS03265	CR922_RS06185	CR921_RS03370	STH8232_RS03650
	shikimate dehydrogenase	*aroE*	BAY21_RS01645	CW339_RS03270	CR922_RS06180	CR921_RS03375	STH8232_RS03655
	shikimate kinase	*aroK*	BAY21_RS01610	CW339_RS03305	CR922_RS06145	CR921_RS03410	STH8232_RS03690
	3-phosphoshikimate 1-carboxyvinyltransferase	*aroA*	BAY21_RS01615	CW339_RS03300	CR922_RS06150	CR921_RS03405	STH8232_RS03685
	chorismate synthase	*aroC*	BAY21_RS01635	CW339_RS03280	CR922_RS06170	CR921_RS03385	STH8232_RS03665
	aminodeoxychorismate/anthranilate synthase component II	*trpG*	BAY21_RS06895	CW339_RS07645	CR922_RS01890	CR921_RS07730	STH8232_RS07905
	anthranilate synthase component I	*trpE*	BAY21_RS06890	CW339_RS07650	CR922_RS01885	CR921_RS07735	STH8232_RS07910
	anthranilate phosphoribosyltransferase	*trpD*	BAY21_RS06900	CW339_RS07640	CR922_RS01895	CR921_RS07725	STH8232_RS07900
	phosphoribosylanthranilate isomerase	*trpF*	BAY21_RS06910	CW339_RS07630	CR922_RS01905	CR921_RS07715	STH8232_RS07890
	indole-3-glycerol phosphate synthase TrpC	*trpC*	BAY21_RS06905	CW339_RS07635	CR922_RS01900	CR921_RS07720	STH8232_RS07895
	tryptophan synthase subunit alpha	*trpA*	BAY21_RS06920	CW339_RS07620	CR922_RS01915	CR921_RS07705	STH8232_RS07880
	tryptophan synthase subunit beta	*trpB*	BAY21_RS06915	CW339_RS07625	CR922_RS01910	CR921_RS07710	STH8232_RS07885
Phenylalanine biosynthesis	chorismate mutase	*aroH*	BAY21_RS08785	CW339_RS05660	CR922_RS03835	CR921_RS05670	STH8232_RS05965
	prephenate dehydratase	*pheA2*	BAY21_RS01605	CW339_RS03310	CR922_RS06140	CR921_RS03415	STH8232_RS03695
	pyridoxal phosphate-dependent aminotransferase		BAY21_RS00815	CW339_RS04190	CR922_RS05335	CR921_RS04135	STH8232_RS04455
Tyrosine biosynthesis	chorismate mutase	*aroH*	BAY21_RS08785	CW339_RS05660	CR922_RS03835	CR921_RS05670	STH8232_RS05965
	prephenate dehydrogenase	*tyrA*	BAY21_RS01630	CW339_RS03285	CR922_RS06165	CR921_RS03390	STH8232_RS03670

	pyridoxal phosphate-dependent aminotransferase		BAY21_RS00815	CW339_RS04190	CR922_RS05335	CR921_RS04135	STH8232_RS04455
Alanine biosynthesis	pyridoxal phosphate-dependent aminotransferase		BAY21_RS06680	CW339_RS07870	CR922_RS01665	CR921_RS07975	STH8232_RS08120
Aspartate biosynthesis	pyridoxal phosphate-dependent aminotransferase		BAY21_RS00815	CW339_RS04190	CR922_RS05335	CR921_RS04135	STH8232_RS04455
Glutamate biosynthesis	pyridoxal phosphate-dependent aminotransferase		BAY21_RS00815	CW339_RS04190	CR922_RS05335	CR921_RS04135	STH8232_RS04455
	pyridoxal phosphate-dependent aminotransferase		BAY21_RS06680	CW339_RS07870	CR922_RS01665	CR921_RS07975	STH8232_RS08120
Lysine biosynthesis (succinyl-Dap pathway)	aspartate kinase	*lysC*	BAY21_RS02875	CW339_RS01930	CR922_RS07425	CR921_RS01995	STH8232_RS02160
	aspartate-semialdehyde dehydrogenase	*asd*	BAY21_RS08220	CW339_RS06305	CR922_RS03265	CR921_RS06300	STH8232_RS06600
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	BAY21_RS08225	CW339_RS06300	CR922_RS03270	CR921_RS06295	STH8232_RS06595
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	BAY21_RS02685	CW339_RS02110	CR922_RS07230	CR921_RS02225	STH8232_RS02340
	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase	*dapD*	BAY21_RS05735	CW339_RS08840	CR922_RS08780	CR921_RS08845	STH8232_RS09095
	acetylornithine transaminase	*argD*	BAY21_RS02480	CW339_RS02320	CR922_RS07025	CR921_RS02430	STH8232_RS02545
	succinyl-diaminopimelate desuccinylase	*dapE*	-	-	-	-	-
	diaminopimelate epimerase	*dapF*	-	-	-	-	-
	diaminopimelate decarboxylase	*lysA*	BAY21_RS03475	CW339_RS01295	CR922_RS08035	CR921_RS01370	STH8232_RS01530
	aspartate kinase	*lysC*	BAY21_RS02875	CW339_RS01930	CR922_RS07425	CR921_RS01995	STH8232_RS02160

Lysine biosynthesis (acetyl-Dap pathway)	aspartate-semialdehyde dehydrogenase	*asd*	BAY21_RS08220	CW339_RS06305	CR922_RS03265	CR921_RS06300	STH8232_RS06600
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	BAY21_RS08225	CW339_RS06300	CR922_RS03270	CR921_RS06295	STH8232_RS06595
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	BAY21_RS02685	CW339_RS02110	CR922_RS07230	CR921_RS02225	STH8232_RS02340
	tetrahydrodipicolinate N-acetyltransferase	*dapH*	-	-	-	-	-
	pyridoxal phosphate-dependent aminotransferase	*patA*	BAY21_RS04615	CW339_RS00200	CR922_RS00210	CR921_RS00215	STH8232_RS00220
	N-acetyldiaminopimelate deacetylase	*dapL*	BAY21_RS05740	CW339_RS08835	CR922_RS08775	CR921_RS08840	STH8232_RS09090
	diaminopimelate epimerase	*dapF*	-	-	-	-	-
	diaminopimelate decarboxylase	*lysA*	BAY21_RS03475	CW339_RS01295	CR922_RS08035	CR921_RS01370	STH8232_RS01530

amino acid	**product**	**gene**	**KLDS 3.1003**	**KLDS SM**	**LMD-9**	**LMG 18311**	**MN-BM-A01**
Serine biosynthesis	3-phosphoglycerate dehydrogenase	*serA*	BEN15_RS00670	A9497_RS03535	STER_RS07310	STU_RS16500	AMD33_RS06220
	3-phosphoserine/phosphohydroxythreonine transaminase	*serC*	BEN15_RS00680	A9497_RS03545	STER_RS07320	STU_RS16510	AMD33_RS06230
	phosphoserine phosphatase SerB	*serB*	BEN15_RS00625	A9497_RS03490	STER_RS07265	STU_RS16455	AMD33_RS06175
Threonine biosynthesis	aspartate kinase	*lysC*	BEN15_RS05060	A9497_RS08045	STER_RS02080	STU_RS11520	AMD33_RS00945
	aspartate-semialdehyde dehydrogenase	*asd*	BEN15_RS09555	A9497_RS02545	STER_RS06290	STU_RS15530	AMD33_RS05215
	homoserine dehydrogenase	*hom*	BEN15_RS05525	A9497_RS08435	STER_RS02485	STU_RS11915	AMD33_RS01370
	homoserine kinase	*thrB*	BEN15_RS05530	A9497_RS08440	STER_RS02490	STU_RS11920	AMD33_RS01375
	threonine synthase	*thrC*	BEN15_RS02320	A9497_RS05250	STER_RS09070	STU_RS18215	AMD33_RS07925
Cysteine biosynthesis	serine O-acetyltransferase	*cysE*	BEN15_RS03615	A9497_RS06580	STER_RS00580	STU_RS10045	AMD33_RS09240

	serine acetyltransferase	*cysE*	BEN15_RS07520	A9497_RS09840	STER_RS04390	STU_RS13595	AMD33_RS04185
	serine acetyltransferase	*cysE*	-	-	-	STU_RS14610	-
	cysteine synthase A	*cysK*	BEN15_RS04955	A9497_RS07940	STER_RS01975	STU_RS11415	AMD33_RS00835
	cysteine synthase family protein	*cysK*	BEN15_RS07510	A9497_RS00635	STER_RS04380	STU_RS13585	AMD33_RS04195
Methionine biosynthesis	aspartate kinase	*lysC*	BEN15_RS05060	A9497_RS08045	STER_RS02080	STU_RS11520	AMD33_RS00945
	aspartate-semialdehyde dehydrogenase	*asd*	BEN15_RS09555	A9497_RS02545	STER_RS06290	STU_RS15530	AMD33_RS05215
	homoserine dehydrogenase	*hom*	BEN15_RS05525	A9497_RS08435	STER_RS02485	STU_RS11915	AMD33_RS01370
	homoserine O-succinyltransferase	*metA*	BEN15_RS09120	A9497_RS02100	STER_RS05865	STU_RS15165	AMD33_RS04765
	cystathionine gamma-synthase	*metB*	BEN15_RS04880	A9497_RS07865	STER_RS01900	STU_RS11340	AMD33_RS00765
	aminotransferase class V-fold PLP-dependent enzyme	*metC*	BEN15_RS07515	A9497_RS00640	STER_RS04385	STU_RS13590	AMD33_RS04190
	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	*metE*	BEN15_RS07180	A9497_RS00335	STER_RS04080	STU_RS13335	AMD33_RS02885
Histidine biosynthesis	ATP phosphoribosyltransferase	*hisG*	BEN15_RS09195	A9497_RS02180	STER_RS05945	-	AMD33_RS04840
	ATP phosphoribosyltransferase regulatory subunit	*hisZ*	BEN15_RS09200	A9497_RS02185	STER_RS05950	-	AMD33_RS04845
	phosphoribosyl-ATP diphosphatase	*hisE*	BEN15_RS09160	A9497_RS02145	STER_RS05910	-	AMD33_RS04805
	phosphoribosyl-AMP cyclohydrolase	*hisI*	BEN15_RS09165	A9497_RS02150	STER_RS05915	-	AMD33_RS04810
	1-(5-phosphoribosyl)-5-((5-phosphoribosylamino)methylideneamino)imidazole-4- carboxamide isomerase	*hisA*	BEN15_RS09175	A9497_RS02160	STER_RS05925	-	AMD33_RS04820

	imidazole glycerol phosphate synthase subunit HisH	*hisH*	BEN15_RS09180	A9497_RS02165	STER_RS05930	-	AMD33_RS04825
	imidazole glycerol phosphate synthase subunit HisF	*hisF*	BEN15_RS09170	A9497_RS02155	STER_RS05920	-	AMD33_RS04815
	imidazoleglycerol-phosphate dehydratase	*hisB*	BEN15_RS09185	A9497_RS02170	STER_RS05935	-	AMD33_RS04830
	histidinol-phosphate transaminase	*hisC*	BEN15_RS09205	A9497_RS02190	STER_RS05955	-	AMD33_RS04850
	hypothetical protein	*hisK*	BEN15_RS09250	A9497_RS02220	STER_RS05985	STU_RS15215	AMD33_RS04890
	histidinol dehydrogenase	*hisD*	BEN15_RS09190	A9497_RS02175	STER_RS05940	-	AMD33_RS04835
Glycine biosynthesis	serine hydroxymethyltransferase	*glyA*	BEN15_RS07105	A9497_RS00185	STER_RS03905	STU_RS13250	AMD33_RS02725
Proline biosynthesis	glutamate-5-semialdehyde dehydrogenase	*proA*	BEN15_RS01495	A9497_RS04410	STER_RS08190	STU_RS17385	AMD33_RS07085
	glutamate 5-kinase	*proB*	BEN15_RS01500	A9497_RS04415	STER_RS08195	STU_RS17390	AMD33_RS07090
	pyrroline-5-carboxylate reductase	*proC*	BEN15_RS02185	A9497_RS05115	STER_RS08930	STU_RS18085	AMD33_RS07790
Glutamine biosynthesis	type I glutamate--ammonia ligase	*glnA*	BEN15_RS01810	A9497_RS04725	STER_RS08555	STU_RS17705	AMD33_RS07400
Arginine biosynthesis	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	BEN15_RS05505	A9497_RS08415	STER_RS02460	STU_RS11895	AMD33_RS01350
	acetylglutamate kinase	*argB*	BEN15_RS05510	A9497_RS08420	STER_RS02465	STU_RS11900	AMD33_RS01355
	N-acetyl-gamma-glutamyl-phosphate reductase	*argC*	BEN15_RS05500	A9497_RS08410	STER_RS02455	STU_RS11890	AMD33_RS01345
	acetylornithine transaminase	*argD*	BEN15_RS05515	A9497_RS08425	STER_RS02470	STU_RS11905	AMD33_RS01360
	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	BEN15_RS05505	A9497_RS08415	STER_RS02460	STU_RS11895	AMD33_RS01350

	ornithine carbamoyltransferase	*argF*	BEN15_RS06175	A9497_RS09165	STER_RS03200	STU_RS12570	AMD33_RS02020
	argininosuccinate synthase	*argG*	BEN15_RS02010	A9497_RS04935	STER_RS08765	STU_RS17905	AMD33_RS07610
	argininosuccinate lyase	*argH*	BEN15_RS02005	A9497_RS04930	STER_RS08760	STU_RS17900	AMD33_RS07605
Asparagine biosynthesis (from aspartate)	asparagine synthetase A	*asnA*	BEN15_RS05055	A9497_RS08040	STER_RS02075	STU_RS11515	AMD33_RS00940
Valine biosynthesis	acetolactate synthase AlsS	*ilvB*	BEN15_RS07760	A9497_RS00905	STER_RS04645	STU_RS13905	AMD33_RS03935
	acetolactate synthase I large subunit	*ilvB*	BEN15_RS02295	A9497_RS05225	STER_RS09040	STU_RS18190	AMD33_RS07900
	acetolactate synthase small subunit	*ilvN*	BEN15_RS02290	A9497_RS05220	STER_RS09035	STU_RS18185	AMD33_RS07895
	ketol-acid reductoisomerase	*ilvC*	BEN15_RS02285	A9497_RS05215	STER_RS09030	STU_RS18180	AMD33_RS07890
	dihydroxy-acid dehydratase	*ilvD*	BEN15_RS01075	A9497_RS03910	STER_RS07700	STU_RS16885	AMD33_RS06600
	dihydroxy-acid dehydratase	*ilvD*	BEN15_RS02300	A9497_RS05230	STER_RS09045	STU_RS18195	AMD33_RS07905
	branched-chain amino acid aminotransferase	*ilvE*	BEN15_RS06095	A9497_RS09090	STER_RS03115	STU_RS12485	AMD33_RS01945
Isoleucine biosynthesis	threonine dehydratase	*ilvA*	BEN15_RS03965	A9497_RS06960	STER_RS01005	STU_RS10440	AMD33_RS09640
	acetolactate synthase AlsS	*ilvB*	BEN15_RS07760	A9497_RS00905	STER_RS04645	STU_RS13905	AMD33_RS03935
	acetolactate synthase I large subunit	*ilvB*	BEN15_RS02295	A9497_RS05225	STER_RS09040	STU_RS18190	AMD33_RS07900
	acetolactate synthase small subunit	*ilvN*	BEN15_RS02290	A9497_RS05220	STER_RS09035	STU_RS18185	AMD33_RS07895
	ketol-acid reductoisomerase	*ilvC*	BEN15_RS02285	A9497_RS05215	STER_RS09030	STU_RS18180	AMD33_RS07890
	dihydroxy-acid dehydratase	*ilvD*	BEN15_RS01075	A9497_RS03910	STER_RS07700	STU_RS16885	AMD33_RS06600
	dihydroxy-acid dehydratase	*ilvD*	BEN15_RS02300	A9497_RS05230	STER_RS09045	STU_RS18195	AMD33_RS07905
	branched-chain amino acid aminotransferase	*ilvE*	BEN15_RS06095	A9497_RS09090	STER_RS03115	STU_RS12485	AMD33_RS01945
Leucine biosynthesis	acetolactate synthase AlsS	*ilvB*	BEN15_RS07760	A9497_RS00905	STER_RS04645	STU_RS13905	AMD33_RS03935

	acetolactate synthase I large subunit	*ilvB*	BEN15_RS02295	A9497_RS05225	STER_RS09040	STU_RS18190	AMD33_RS07900
	acetolactate synthase small subunit	*ilvN*	BEN15_RS02290	A9497_RS05220	STER_RS09035	STU_RS18185	AMD33_RS07895
	ketol-acid reductoisomerase	*ilvC*	BEN15_RS02285	A9497_RS05215	STER_RS09030	STU_RS18180	AMD33_RS07890
	dihydroxy-acid dehydratase	*ilvD*	BEN15_RS01075	A9497_RS03910	STER_RS07700	STU_RS16885	AMD33_RS06600
	2-isopropylmalate synthase	*leuA*	BEN15_RS09035	A9497_RS02015	STER_RS05780	STU_RS15080	AMD33_RS04680
	3-isopropylmalate dehydratase large subunit	*leuC*	BEN15_RS09020	A9497_RS02000	STER_RS05765	STU_RS15065	AMD33_RS04665
	3-isopropylmalate dehydratase small subunit	*leuD*	BEN15_RS09015	A9497_RS01995	STER_RS05760	STU_RS15060	AMD33_RS04660
	3-isopropylmalate dehydrogenase	*leuB*	BEN15_RS09030	A9497_RS02010	STER_RS05775	STU_RS15075	AMD33_RS04675
	branched-chain amino acid aminotransferase	*ilvE*	BEN15_RS06095	A9497_RS09090	STER_RS03115	STU_RS12485	AMD33_RS01945
Tryptophan biosynthesis	3-deoxy-7-phosphoheptulonate synthase	*aroG1*	BEN15_RS01580	A9497_RS04495	STER_RS08325	STU_RS17470	AMD33_RS07170
	3-deoxy-7-phosphoheptulonate synthase	*aroG2*	BEN15_RS01585	A9497_RS04500	STER_RS08330	STU_RS17475	AMD33_RS07175
	3-dehydroquinate synthase	*aroB*	BEN15_RS06365	A9497_RS09355	STER_RS03395	STU_RS12750	AMD33_RS02210
	type I 3-dehydroquinate dehydratase	*aroD*	BEN15_RS06355	A9497_RS09345	STER_RS03385	STU_RS12740	AMD33_RS02200
	shikimate dehydrogenase	*aroE*	BEN15_RS06360	A9497_RS09350	STER_RS03390	STU_RS12745	AMD33_RS02205
	shikimate kinase	*aroK*	BEN15_RS06395	A9497_RS09385	STER_RS03425	STU_RS12780	AMD33_RS02240
	3-phosphoshikimate 1-carboxyvinyltransferase	*aroA*	BEN15_RS06390	A9497_RS09380	STER_RS03420	STU_RS12775	AMD33_RS02235
	chorismate synthase	*aroC*	BEN15_RS06370	A9497_RS09360	STER_RS03400	STU_RS12755	AMD33_RS02215
	aminodeoxychorismate/anthranilate synthase component II	*trpG*	BEN15_RS00980	A9497_RS03850	STER_RS07635	STU_RS16825	AMD33_RS06540

	anthranilate synthase component I	*trpE*	BEN15_RS00985	A9497_RS03855	STER_RS07640	STU_RS16830	AMD33_RS06545
	anthranilate phosphoribosyltransferase	*trpD*	BEN15_RS00975	A9497_RS03845	STER_RS07630	STU_RS16820	AMD33_RS06535
	phosphoribosylanthranilate isomerase	*trpF*	BEN15_RS00965	A9497_RS03835	STER_RS07620	STU_RS16810	AMD33_RS06525
	indole-3-glycerol phosphate synthase TrpC	*trpC*	BEN15_RS00970	A9497_RS03840	STER_RS07625	STU_RS16815	AMD33_RS06530
	tryptophan synthase subunit alpha	*trpA*	BEN15_RS00955	A9497_RS03825	STER_RS07610	STU_RS16800	AMD33_RS06515
	tryptophan synthase subunit beta	*trpB*	BEN15_RS00960	A9497_RS03830	STER_RS07615	STU_RS16805	AMD33_RS06520
Phenylalanine biosynthesis	chorismate mutase	*aroH*	BEN15_RS08920	A9497_RS01890	STER_RS05655	STU_RS14960	AMD33_RS04550
	prephenate dehydratase	*pheA2*	BEN15_RS06400	A9497_RS09390	STER_RS03430	STU_RS12785	AMD33_RS02245
	pyridoxal phosphate-dependent aminotransferase		BEN15_RS07300	A9497_RS00515	STER_RS04255	STU_RS13470	AMD33_RS04315
Tyrosine biosynthesis	chorismate mutase	*aroH*	BEN15_RS08920	A9497_RS01890	STER_RS05655	STU_RS14960	AMD33_RS04550
	prephenate dehydrogenase	*tyrA*	BEN15_RS06375	A9497_RS09365	STER_RS03405	STU_RS12760	AMD33_RS02220
	pyridoxal phosphate-dependent aminotransferase		BEN15_RS07300	A9497_RS00515	STER_RS04255	STU_RS13470	AMD33_RS04315
Alanine biosynthesis	pyridoxal phosphate-dependent aminotransferase		BEN15_RS01230	A9497_RS04065	STER_RS07855	STU_RS17040	AMD33_RS06755
Aspartate biosynthesis	pyridoxal phosphate-dependent aminotransferase		BEN15_RS07300	A9497_RS00515	STER_RS04255	STU_RS13470	AMD33_RS04315
Glutamate biosynthesis	pyridoxal phosphate-dependent aminotransferase		BEN15_RS07300	A9497_RS00515	STER_RS04255	STU_RS13470	AMD33_RS04315
	pyridoxal phosphate-dependent aminotransferase		BEN15_RS01230	A9497_RS04065	STER_RS07855	STU_RS17040	AMD33_RS06755
	aspartate kinase	*lysC*	BEN15_RS05060	A9497_RS08045	STER_RS02080	STU_RS11520	AMD33_RS00945

Lysine biosynthesis (succinyl-Dap pathway)	aspartate-semialdehyde dehydrogenase	*asd*	BEN15_RS09555	A9497_RS02545	STER_RS06290	STU_RS15530	AMD33_RS05215
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	BEN15_RS09550	A9497_RS02540	STER_RS06285	STU_RS15525	AMD33_RS05210
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	BEN15_RS05315	A9497_RS08225	STER_RS02260	STU_RS11710	AMD33_RS01165
	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase	*dapD*	BEN15_RS02115	A9497_RS05045	STER_RS08865	STU_RS18020	AMD33_RS07720
	acetylornithine transaminase	*argD*	BEN15_RS05515	A9497_RS08425	STER_RS02470	STU_RS11905	AMD33_RS01360
	succinyl-diaminopimelate desuccinylase	*dapE*	-	-	-	-	-
	diaminopimelate epimerase	*dapF*	-	-	-	-	-
	diaminopimelate decarboxylase	*lysA*	BEN15_RS04415	A9497_RS07415	STER_RS01450	STU_RS10890	AMD33_RS00320
Lysine biosynthesis (acetyl-Dap pathway)	aspartate kinase	*lysC*	BEN15_RS05060	A9497_RS08045	STER_RS02080	STU_RS11520	AMD33_RS00945
	aspartate-semialdehyde dehydrogenase	*asd*	BEN15_RS09555	A9497_RS02545	STER_RS06290	STU_RS15530	AMD33_RS05215
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	BEN15_RS09550	A9497_RS02540	STER_RS06285	STU_RS15525	AMD33_RS05210
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	BEN15_RS05315	A9497_RS08225	STER_RS02260	STU_RS11710	AMD33_RS01165
	tetrahydrodipicolinate N-acetyltransferase	*dapH*	-	-	-	-	-
	pyridoxal phosphate-dependent aminotransferase	*patA*	BEN15_RS03270	A9497_RS06230	STER_RS00230	STU_RS09685	AMD33_RS08900
	N-acetyldiaminopimelate deacetylase	*dapL*	BEN15_RS02110	A9497_RS05040	STER_RS08860	STU_RS18015	AMD33_RS07715

	diaminopimelate epimerase	*dapF*	-	-	-	-	-
	diaminopimelate decarboxylase	*lysA*	BEN15_RS04415	A9497_RS07415	STER_RS01450	STU_RS10890	AMD33_RS00320

amino acid	**product**	**gene**	**MN-BM-A02**	**MN-ZLW-002**	**NCTC12958**	**ND03**	**ND07**
Serine biosynthesis	3-phosphoglycerate dehydrogenase	*serA*	MNA02_RS07225	Y1U_RS07240	DQL34_RS08375	STND_RS07230	BBD27_RS06815
	3-phosphoserine/phosphohydroxythreonine transaminase	*serC*	MNA02_RS07235	Y1U_RS07250	DQL34_RS08385	STND_RS07240	BBD27_RS06805
	phosphoserine phosphatase SerB	*serB*	MNA02_RS07180	Y1U_RS07195	DQL34_RS08330	STND_RS07185	BBD27_RS06860
Threonine biosynthesis	aspartate kinase	*lysC*	MNA02_RS01960	Y1U_RS01985	DQL34_RS02330	STND_RS02020	BBD27_RS02390
	aspartate-semialdehyde dehydrogenase	*asd*	MNA02_RS06220	Y1U_RS06225	DQL34_RS07240	STND_RS06220	BBD27_RS07805
	homoserine dehydrogenase	*hom*	MNA02_RS02350	Y1U_RS02410	DQL34_RS02860	STND_RS02465	BBD27_RS02000
	homoserine kinase	*thrB*	MNA02_RS02355	Y1U_RS02415	DQL34_RS02865	STND_RS02470	BBD27_RS01995
	threonine synthase	*thrC*	MNA02_RS08945	Y1U_RS08935	DQL34_RS10405	STND_RS08930	BBD27_RS05095
Cysteine biosynthesis	serine O-acetyltransferase	*cysE*	MNA02_RS00570	Y1U_RS00560	DQL34_RS00615	STND_RS00600	BBD27_RS03770
	serine acetyltransferase	*cysE*	MNA02_RS10155	Y1U_RS05200	-	STND_RS04245	BBD27_RS09720
	serine acetyltransferase	*cysE*	-	-	-	-	-
	cysteine synthase A	*cysK*	MNA02_RS01855	Y1U_RS01880	DQL34_RS02220	STND_RS01915	BBD27_RS02495
	cysteine synthase family protein	*cysK*	MNA02_RS04295	Y1U_RS05210	-	STND_RS04235	BBD27_RS09730
Methionine biosynthesis	aspartate kinase	*lysC*	MNA02_RS01960	Y1U_RS01985	DQL34_RS02330	STND_RS02020	BBD27_RS02390
	aspartate-semialdehyde dehydrogenase	*asd*	MNA02_RS06220	Y1U_RS06225	DQL34_RS07240	STND_RS06220	BBD27_RS07805
	homoserine dehydrogenase	*hom*	MNA02_RS02350	Y1U_RS02410	DQL34_RS02860	STND_RS02465	BBD27_RS02000
	homoserine O-succinyltransferase	*metA*	MNA02_RS05770	Y1U_RS05775	DQL34_RS06805	STND_RS05780	BBD27_RS08250
	cystathionine gamma-synthase	*metB*	MNA02_RS01780	Y1U_RS01810	DQL34_RS02150	STND_RS01845	BBD27_RS02570

	aminotransferase class V-fold PLP-dependent enzyme	*metC*	MNA02_RS04300	Y1U_RS05205	-	STND_RS04240	BBD27_RS09725
	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	*metE*	MNA02_RS03990	Y1U_RS03925	DQL34_RS04875	STND_RS03975	BBD27_RS00350
Histidine biosynthesis	ATP phosphoribosyltransferase	*hisG*	MNA02_RS05850	Y1U_RS05850	DQL34_RS06880	STND_RS05855	BBD27_RS08170
	ATP phosphoribosyltransferase regulatory subunit	*hisZ*	MNA02_RS05855	Y1U_RS05855	DQL34_RS06885	STND_RS05860	BBD27_RS08165
	phosphoribosyl-ATP diphosphatase	*hisE*	MNA02_RS05815	Y1U_RS05815	DQL34_RS06845	STND_RS05820	BBD27_RS08205
	phosphoribosyl-AMP cyclohydrolase	*hisI*	MNA02_RS05820	Y1U_RS05820	DQL34_RS06850	STND_RS05825	BBD27_RS08200
	1-(5-phosphoribosyl)-5-((5-phosphoribosylamino)methylideneamino)imidazole-4- carboxamide isomerase	*hisA*	MNA02_RS05830	Y1U_RS05830	DQL34_RS06860	STND_RS05835	BBD27_RS08190
	imidazole glycerol phosphate synthase subunit HisH	*hisH*	MNA02_RS05835	Y1U_RS05835	DQL34_RS06865	STND_RS05840	BBD27_RS08185
	imidazole glycerol phosphate synthase subunit HisF	*hisF*	MNA02_RS05825	Y1U_RS05825	DQL34_RS06855	STND_RS05830	BBD27_RS08195
	imidazoleglycerol-phosphate dehydratase	*hisB*	MNA02_RS05840	Y1U_RS05840	DQL34_RS06870	STND_RS05845	BBD27_RS08180
	histidinol-phosphate transaminase	*hisC*	MNA02_RS05860	Y1U_RS05860	DQL34_RS06890	STND_RS05865	BBD27_RS08160
	hypothetical protein	*hisK*	MNA02_RS05890	Y1U_RS05900	DQL34_RS06910	STND_RS05900	BBD27_RS08130
	histidinol dehydrogenase	*hisD*	MNA02_RS05845	Y1U_RS05845	DQL34_RS06875	STND_RS05850	BBD27_RS08175

Glycine biosynthesis	serine hydroxymethyltransferase	*glyA*	MNA02_RS03835	Y1U_RS03765	DQL34_RS04805	STND_RS03815	BBD27_RS00490
Proline biosynthesis	glutamate-5-semialdehyde dehydrogenase	*proA*	MNA02_RS08110	Y1U_RS08105	DQL34_RS09450	STND_RS08090	BBD27_RS05935
	glutamate 5-kinase	*proB*	MNA02_RS08115	Y1U_RS08110	DQL34_RS09455	STND_RS08095	BBD27_RS05930
	pyrroline-5-carboxylate reductase	*proC*	MNA02_RS08810	Y1U_RS08805	DQL34_RS10275	STND_RS08795	BBD27_RS05230
Glutamine biosynthesis	type I glutamate--ammonia ligase	*glnA*	MNA02_RS08425	Y1U_RS08420	DQL34_RS09900	STND_RS08405	BBD27_RS05620
Arginine biosynthesis	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	MNA02_RS02330	Y1U_RS02390	DQL34_RS02840	STND_RS02445	BBD27_RS02020
	acetylglutamate kinase	*argB*	MNA02_RS02335	Y1U_RS02395	DQL34_RS02845	STND_RS02450	BBD27_RS02015
	N-acetyl-gamma-glutamyl-phosphate reductase	*argC*	MNA02_RS02325	Y1U_RS02385	DQL34_RS02835	STND_RS02440	BBD27_RS02025
	acetylornithine transaminase	*argD*	MNA02_RS02340	Y1U_RS02400	DQL34_RS02850	STND_RS02455	BBD27_RS02010
	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	MNA02_RS02330	Y1U_RS02390	DQL34_RS02840	STND_RS02445	BBD27_RS02020
	ornithine carbamoyltransferase	*argF*	MNA02_RS03060	Y1U_RS03060	DQL34_RS03720	STND_RS03135	BBD27_RS01265
	argininosuccinate synthase	*argG*	MNA02_RS08640	Y1U_RS08635	DQL34_RS10105	STND_RS08615	BBD27_RS05410
	argininosuccinate lyase	*argH*	MNA02_RS08635	Y1U_RS08630	DQL34_RS10100	STND_RS08610	BBD27_RS05415
Asparagine biosynthesis (from aspartate)	asparagine synthetase A	*asnA*	MNA02_RS01955	Y1U_RS01980	DQL34_RS02325	STND_RS02015	BBD27_RS02395
Valine biosynthesis	acetolactate synthase AlsS	*ilvB*	MNA02_RS04580	Y1U_RS04950	DQL34_RS05680	STND_RS04485	BBD27_RS09445
	acetolactate synthase I large subunit	*ilvB*	MNA02_RS08920	Y1U_RS08910	DQL34_RS10380	STND_RS08905	BBD27_RS05120
	acetolactate synthase small subunit	*ilvN*	MNA02_RS08915	Y1U_RS08905	DQL34_RS10375	STND_RS08900	BBD27_RS05125

	ketol-acid reductoisomerase	*ilvC*	MNA02_RS08910	Y1U_RS08900	DQL34_RS10370	STND_RS08895	BBD27_RS05130
	dihydroxy-acid dehydratase	*ilvD*	MNA02_RS07605	Y1U_RS07620	DQL34_RS08820	STND_RS07610	BBD27_RS06435
	dihydroxy-acid dehydratase	*ilvD*	MNA02_RS08925	Y1U_RS08915	DQL34_RS10385	STND_RS08910	BBD27_RS05115
	branched-chain amino acid aminotransferase	*ilvE*	MNA02_RS02990	Y1U_RS02990	DQL34_RS03650	STND_RS03060	BBD27_RS01340
Isoleucine biosynthesis	threonine dehydratase	*ilvA*	MNA02_RS00885	Y1U_RS00895	DQL34_RS01230	STND_RS00940	BBD27_RS03460
	acetolactate synthase AlsS	*ilvB*	MNA02_RS04580	Y1U_RS04950	DQL34_RS05680	STND_RS04485	BBD27_RS09445
	acetolactate synthase I large subunit	*ilvB*	MNA02_RS08920	Y1U_RS08910	DQL34_RS10380	STND_RS08905	BBD27_RS05120
	acetolactate synthase small subunit	*ilvN*	MNA02_RS08915	Y1U_RS08905	DQL34_RS10375	STND_RS08900	BBD27_RS05125
	ketol-acid reductoisomerase	*ilvC*	MNA02_RS08910	Y1U_RS08900	DQL34_RS10370	STND_RS08895	BBD27_RS05130
	dihydroxy-acid dehydratase	*ilvD*	MNA02_RS07605	Y1U_RS07620	DQL34_RS08820	STND_RS07610	BBD27_RS06435
	dihydroxy-acid dehydratase	*ilvD*	MNA02_RS08925	Y1U_RS08915	DQL34_RS10385	STND_RS08910	BBD27_RS05115
	branched-chain amino acid aminotransferase	*ilvE*	MNA02_RS02990	Y1U_RS02990	DQL34_RS03650	STND_RS03060	BBD27_RS01340
Leucine biosynthesis	acetolactate synthase AlsS	*ilvB*	MNA02_RS04580	Y1U_RS04950	DQL34_RS05680	STND_RS04485	BBD27_RS09445
	acetolactate synthase I large subunit	*ilvB*	MNA02_RS08920	Y1U_RS08910	DQL34_RS10380	STND_RS08905	BBD27_RS05120
	acetolactate synthase small subunit	*ilvN*	MNA02_RS08915	Y1U_RS08905	DQL34_RS10375	STND_RS08900	BBD27_RS05125
	ketol-acid reductoisomerase	*ilvC*	MNA02_RS08910	Y1U_RS08900	DQL34_RS10370	STND_RS08895	BBD27_RS05130
	dihydroxy-acid dehydratase	*ilvD*	MNA02_RS07605	Y1U_RS07620	DQL34_RS08820	STND_RS07610	BBD27_RS06435
	2-isopropylmalate synthase	*leuA*	MNA02_RS05685	Y1U_RS05690	DQL34_RS06720	STND_RS05695	BBD27_RS08335
	3-isopropylmalate dehydratase large subunit	*leuC*	MNA02_RS05670	Y1U_RS05675	DQL34_RS06705	STND_RS05680	BBD27_RS08350

	3-isopropylmalate dehydratase small subunit	*leuD*	MNA02_RS05665	Y1U_RS05670	DQL34_RS06700	STND_RS05675	BBD27_RS08355
	3-isopropylmalate dehydrogenase	*leuB*	MNA02_RS05680	Y1U_RS05685	DQL34_RS06715	STND_RS05690	BBD27_RS08340
	branched-chain amino acid aminotransferase	*ilvE*	MNA02_RS02990	Y1U_RS02990	DQL34_RS03650	STND_RS03060	BBD27_RS01340
Tryptophan biosynthesis	3-deoxy-7-phosphoheptulonate synthase	*aroG1*	MNA02_RS08195	Y1U_RS08190	DQL34_RS09535	STND_RS08175	BBD27_RS05850
	3-deoxy-7-phosphoheptulonate synthase	*aroG2*	MNA02_RS08200	Y1U_RS08195	DQL34_RS09540	STND_RS08180	BBD27_RS05845
	3-dehydroquinate synthase	*aroB*	MNA02_RS03255	Y1U_RS03255	DQL34_RS04190	STND_RS03330	BBD27_RS01075
	type I 3-dehydroquinate dehydratase	*aroD*	MNA02_RS03245	Y1U_RS03245	DQL34_RS04180	STND_RS03320	BBD27_RS01085
	shikimate dehydrogenase	*aroE*	MNA02_RS03250	Y1U_RS03250	DQL34_RS04185	STND_RS03325	BBD27_RS01080
	shikimate kinase	*aroK*	MNA02_RS03285	Y1U_RS03285	DQL34_RS04220	STND_RS03360	BBD27_RS01045
	3-phosphoshikimate 1-carboxyvinyltransferase	*aroA*	MNA02_RS03280	Y1U_RS03280	DQL34_RS04215	STND_RS03355	BBD27_RS01050
	chorismate synthase	*aroC*	MNA02_RS03260	Y1U_RS03260	DQL34_RS04195	STND_RS03335	BBD27_RS01070
	aminodeoxychorismate/anthranilate synthase component II	*trpG*	MNA02_RS07545	Y1U_RS07560	DQL34_RS08690	STND_RS07550	BBD27_RS06495
	anthranilate synthase component I	*trpE*	MNA02_RS07550	Y1U_RS07565	DQL34_RS08695	STND_RS07555	BBD27_RS06490
	anthranilate phosphoribosyltransferase	*trpD*	MNA02_RS07540	Y1U_RS07555	DQL34_RS08685	STND_RS07545	BBD27_RS06500
	phosphoribosylanthranilate isomerase	*trpF*	MNA02_RS07530	Y1U_RS07545	DQL34_RS08675	STND_RS07535	BBD27_RS06510
	indole-3-glycerol phosphate synthase TrpC	*trpC*	MNA02_RS07535	Y1U_RS07550	DQL34_RS08680	STND_RS07540	BBD27_RS06505
	tryptophan synthase subunit alpha	*trpA*	MNA02_RS07520	Y1U_RS07535	DQL34_RS08665	STND_RS07525	BBD27_RS06520

	tryptophan synthase subunit beta	*trpB*	MNA02_RS07525	Y1U_RS07540	DQL34_RS08670	STND_RS07530	BBD27_RS06515
Phenylalanine biosynthesis	chorismate mutase	*aroH*	MNA02_RS05560	Y1U_RS05560	DQL34_RS06600	STND_RS05565	BBD27_RS08460
	prephenate dehydratase	*pheA2*	MNA02_RS03290	Y1U_RS03290	DQL34_RS04225	STND_RS03365	BBD27_RS01040
	pyridoxal phosphate-dependent aminotransferase		MNA02_RS04170	Y1U_RS05330	DQL34_RS05010	STND_RS04110	BBD27_RS00175
Tyrosine biosynthesis	chorismate mutase	*aroH*	MNA02_RS05560	Y1U_RS05560	DQL34_RS06600	STND_RS05565	BBD27_RS08460
	prephenate dehydrogenase	*tyrA*	MNA02_RS03265	Y1U_RS03265	DQL34_RS04200	STND_RS03340	BBD27_RS01065
	pyridoxal phosphate-dependent aminotransferase		MNA02_RS04170	Y1U_RS05330	DQL34_RS05010	STND_RS04110	BBD27_RS00175
Alanine biosynthesis	pyridoxal phosphate-dependent aminotransferase		MNA02_RS07760	Y1U_RS07775	DQL34_RS09000	STND_RS07765	BBD27_RS06280
Aspartate biosynthesis	pyridoxal phosphate-dependent aminotransferase		MNA02_RS04170	Y1U_RS05330	DQL34_RS05010	STND_RS04110	BBD27_RS00175
Glutamate biosynthesis	pyridoxal phosphate-dependent aminotransferase		MNA02_RS04170	Y1U_RS05330	DQL34_RS05010	STND_RS04110	BBD27_RS00175
	pyridoxal phosphate-dependent aminotransferase		MNA02_RS07760	Y1U_RS07775	DQL34_RS09000	STND_RS07765	BBD27_RS06280
Lysine biosynthesis (succinyl-Dap pathway)	aspartate kinase	*lysC*	MNA02_RS01960	Y1U_RS01985	DQL34_RS02330	STND_RS02020	BBD27_RS02390
	aspartate-semialdehyde dehydrogenase	*asd*	MNA02_RS06220	Y1U_RS06225	DQL34_RS07240	STND_RS06220	BBD27_RS07805
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	MNA02_RS06215	Y1U_RS06220	DQL34_RS07235	STND_RS06215	BBD27_RS07810
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	MNA02_RS02140	Y1U_RS02205	DQL34_RS02500	STND_RS02260	BBD27_RS02210
	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase	*dapD*	MNA02_RS08745	Y1U_RS08740	DQL34_RS10200	STND_RS08730	BBD27_RS05300

	acetylornithine transaminase	*argD*	MNA02_RS02340	Y1U_RS02400	DQL34_RS02850	STND_RS02455	BBD27_RS02010
	succinyl-diaminopimelate desuccinylase	*dapE*	-	-	-	-	-
	diaminopimelate epimerase	*dapF*	-	-	-	-	-
	diaminopimelate decarboxylase	*lysA*	MNA02_RS01330	Y1U_RS01365	DQL34_RS01680	STND_RS01395	BBD27_RS03015
Lysine biosynthesis (acetyl-Dap pathway)	aspartate kinase	*lysC*	MNA02_RS01960	Y1U_RS01985	DQL34_RS02330	STND_RS02020	BBD27_RS02390
	aspartate-semialdehyde dehydrogenase	*asd*	MNA02_RS06220	Y1U_RS06225	DQL34_RS07240	STND_RS06220	BBD27_RS07805
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	MNA02_RS06215	Y1U_RS06220	DQL34_RS07235	STND_RS06215	BBD27_RS07810
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	MNA02_RS02140	Y1U_RS02205	DQL34_RS02500	STND_RS02260	BBD27_RS02210
	tetrahydrodipicolinate N-acetyltransferase	*dapH*	-	-	-	-	-
	pyridoxal phosphate-dependent aminotransferase	*patA*	MNA02_RS00220	Y1U_RS00225	DQL34_RS00210	STND_RS00220	BBD27_RS04115
	N-acetyldiaminopimelate deacetylase	*dapL*	MNA02_RS08740	Y1U_RS08735	DQL34_RS10195	STND_RS08725	BBD27_RS05305
	diaminopimelate epimerase	*dapF*	-	-	-	-	-
	diaminopimelate decarboxylase	*lysA*	MNA02_RS01330	Y1U_RS01365	DQL34_RS01680	STND_RS01395	BBD27_RS03015

amino acid	**product**	**gene**	**S9**	**SMQ-301**	**ST3**
Serine biosynthesis	3-phosphoglycerate dehydrogenase	*serA*	AVT04_RS07370	SMQ301_RS07295	BGL51_RS07515
	3-phosphoserine/phosphohydroxythreonine transaminase	*serC*	AVT04_RS07360	SMQ301_RS07305	BGL51_RS07525

	phosphoserine phosphatase SerB	*serB*	AVT04_RS07415	SMQ301_RS07250	BGL51_RS07470
Threonine biosynthesis	aspartate kinase	*lysC*	AVT04_RS02955	SMQ301_RS02095	BGL51_RS02050
	aspartate-semialdehyde dehydrogenase	*asd*	AVT04_RS08380	SMQ301_RS06300	BGL51_RS06490
	homoserine dehydrogenase	*hom*	AVT04_RS02550	SMQ301_RS02485	BGL51_RS02435
	homoserine kinase	*thrB*	AVT04_RS02545	SMQ301_RS02490	BGL51_RS02440
	threonine synthase	*thrC*	AVT04_RS05690	SMQ301_RS09055	BGL51_RS09145
Cysteine biosynthesis	serine O-acetyltransferase	*cysE*	AVT04_RS04435	SMQ301_RS00585	BGL51_RS00545
	serine acetyltransferase	*cysE*	AVT04_RS00815	SMQ301_RS04405	BGL51_RS04445
	serine acetyltransferase	*cysE*	-	-	-
	cysteine synthase A	*cysK*	AVT04_RS03065	SMQ301_RS01990	BGL51_RS01945
	cysteine synthase family protein	*cysK*	AVT04_RS00825	SMQ301_RS04395	BGL51_RS04435
Methionine biosynthesis	aspartate kinase	*lysC*	AVT04_RS02955	SMQ301_RS02095	BGL51_RS02050
	aspartate-semialdehyde dehydrogenase	*asd*	AVT04_RS08380	SMQ301_RS06300	BGL51_RS06490
	homoserine dehydrogenase	*hom*	AVT04_RS02550	SMQ301_RS02485	BGL51_RS02435
	homoserine O-succinyltransferase	*metA*	AVT04_RS08730	SMQ301_RS05875	BGL51_RS06060
	cystathionine gamma-synthase	*metB*	AVT04_RS03135	SMQ301_RS01915	BGL51_RS01875
	aminotransferase class V-fold PLP-dependent enzyme	*metC*	AVT04_RS00820	SMQ301_RS04400	BGL51_RS04440
	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase	*metE*	AVT04_RS01080	SMQ301_RS04090	BGL51_RS04065
Histidine biosynthesis	ATP phosphoribosyltransferase	*hisG*	-	SMQ301_RS05955	BGL51_RS06140

	ATP phosphoribosyltransferase regulatory subunit	*hisZ*	-	SMQ301_RS05960	BGL51_RS06145
	phosphoribosyl-ATP diphosphatase	*hisE*	-	SMQ301_RS05920	BGL51_RS06105
	phosphoribosyl-AMP cyclohydrolase	*hisI*	-	SMQ301_RS05925	BGL51_RS06110
	1-(5-phosphoribosyl)-5-((5-phosphoribosylamino)methylideneamino)imidazole-4- carboxamide isomerase	*hisA*	-	SMQ301_RS05935	BGL51_RS06120
	imidazole glycerol phosphate synthase subunit HisH	*hisH*	-	SMQ301_RS05940	BGL51_RS06125
	imidazole glycerol phosphate synthase subunit HisF	*hisF*	-	SMQ301_RS05930	BGL51_RS06115
	imidazoleglycerol-phosphate dehydratase	*hisB*	-	SMQ301_RS05945	BGL51_RS06130
	histidinol-phosphate transaminase	*hisC*	-	SMQ301_RS05965	BGL51_RS06150
	hypothetical protein	*hisK*	AVT04_RS08680	SMQ301_RS05995	BGL51_RS06170
	histidinol dehydrogenase	*hisD*	-	SMQ301_RS05950	BGL51_RS06135
Glycine biosynthesis	serine hydroxymethyltransferase	*glyA*	AVT04_RS01170	SMQ301_RS03905	BGL51_RS03905
Proline biosynthesis	glutamate-5-semialdehyde dehydrogenase	*proA*	AVT04_RS06515	SMQ301_RS08180	BGL51_RS08345
	glutamate 5-kinase	*proB*	AVT04_RS06510	SMQ301_RS08185	BGL51_RS08350
	pyrroline-5-carboxylate reductase	*proC*	AVT04_RS05820	SMQ301_RS08920	BGL51_RS09010
Glutamine biosynthesis	type I glutamate--ammonia ligase	*glnA*	AVT04_RS06195	SMQ301_RS08545	BGL51_RS08655

Arginine biosynthesis	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	AVT04_RS02570	SMQ301_RS02465	BGL51_RS02415
	acetylglutamate kinase	*argB*	AVT04_RS02565	SMQ301_RS02470	BGL51_RS02420
	N-acetyl-gamma-glutamyl-phosphate reductase	*argC*	AVT04_RS02575	SMQ301_RS02460	BGL51_RS02410
	acetylornithine transaminase	*argD*	AVT04_RS02560	SMQ301_RS02475	BGL51_RS02425
	bifunctional glutamate N-acetyltransferase/amino-acid acetyltransferase ArgJ	*argJ*	AVT04_RS02570	SMQ301_RS02465	BGL51_RS02415
	ornithine carbamoyltransferase	*argF*	AVT04_RS01895	SMQ301_RS03180	BGL51_RS03155
	argininosuccinate synthase	*argG*	AVT04_RS06000	SMQ301_RS08755	BGL51_RS08845
	argininosuccinate lyase	*argH*	AVT04_RS06005	SMQ301_RS08750	BGL51_RS08840
Asparagine biosynthesis (from aspartate)	asparagine synthetase A	*asnA*	AVT04_RS02960	SMQ301_RS02090	BGL51_RS02045
Valine biosynthesis	acetolactate synthase AlsS	*ilvB*	AVT04_RS00495,	SMQ301_RS04625	BGL51_RS04725
	acetolactate synthase I large subunit	*ilvB*	AVT04_RS05715	SMQ301_RS09030	BGL51_RS09120
	acetolactate synthase small subunit	*ilvN*	AVT04_RS05720	SMQ301_RS09025	BGL51_RS09115
	ketol-acid reductoisomerase	*ilvC*	AVT04_RS05725	SMQ301_RS09020	BGL51_RS09110
	dihydroxy-acid dehydratase	*ilvD*	AVT04_RS06995	SMQ301_RS07685	BGL51_RS07905
	dihydroxy-acid dehydratase	*ilvD*	AVT04_RS05710	SMQ301_RS09035	BGL51_RS09125
	branched-chain amino acid aminotransferase	*ilvE*	AVT04_RS01980	SMQ301_RS03095	BGL51_RS03085
Isoleucine biosynthesis	threonine dehydratase	*ilvA*	AVT04_RS04025	SMQ301_RS01005	BGL51_RS00975
	acetolactate synthase AlsS	*ilvB*	AVT04_RS00495,	SMQ301_RS04625	BGL51_RS04725

	acetolactate synthase I large subunit	*ilvB*	AVT04_RS05715	SMQ301_RS09030	BGL51_RS09120
	acetolactate synthase small subunit	*ilvN*	AVT04_RS05720	SMQ301_RS09025	BGL51_RS09115
	ketol-acid reductoisomerase	*ilvC*	AVT04_RS05725	SMQ301_RS09020	BGL51_RS09110
	dihydroxy-acid dehydratase	*ilvD*	AVT04_RS06995	SMQ301_RS07685	BGL51_RS07905
	dihydroxy-acid dehydratase	*ilvD*	AVT04_RS05710	SMQ301_RS09035	BGL51_RS09125
	branched-chain amino acid aminotransferase	*ilvE*	AVT04_RS01980	SMQ301_RS03095	BGL51_RS03085
Leucine biosynthesis	acetolactate synthase AlsS	*ilvB*	AVT04_RS00495,	SMQ301_RS04625	BGL51_RS04725
	acetolactate synthase I large subunit	*ilvB*	AVT04_RS05715	SMQ301_RS09030	BGL51_RS09120
	acetolactate synthase small subunit	*ilvN*	AVT04_RS05720	SMQ301_RS09025	BGL51_RS09115
	ketol-acid reductoisomerase	*ilvC*	AVT04_RS05725	SMQ301_RS09020	BGL51_RS09110
	dihydroxy-acid dehydratase	*ilvD*	AVT04_RS06995	SMQ301_RS07685	BGL51_RS07905
	2-isopropylmalate synthase	*leuA*	AVT04_RS08825	SMQ301_RS05790	BGL51_RS05975
	3-isopropylmalate dehydratase large subunit	*leuC*	AVT04_RS08840	SMQ301_RS05775	BGL51_RS05960
	3-isopropylmalate dehydratase small subunit	*leuD*	AVT04_RS08845	SMQ301_RS05770	BGL51_RS05955
	3-isopropylmalate dehydrogenase	*leuB*	AVT04_RS08830	SMQ301_RS05785	BGL51_RS05970
	branched-chain amino acid aminotransferase	*ilvE*	AVT04_RS01980	SMQ301_RS03095	BGL51_RS03085
Tryptophan biosynthesis	3-deoxy-7-phosphoheptulonate synthase	*aroG1*	AVT04_RS06430	SMQ301_RS08315	BGL51_RS08430

	3-deoxy-7-phosphoheptulonate synthase	*aroG2*	AVT04_RS06425	SMQ301_RS08320	-
	3-dehydroquinate synthase	*aroB*	AVT04_RS01720	SMQ301_RS03380	BGL51_RS03365
	type I 3-dehydroquinate dehydratase	*aroD*	AVT04_RS01730	SMQ301_RS03370	BGL51_RS03355
	shikimate dehydrogenase	*aroE*	AVT04_RS01725	SMQ301_RS03375	BGL51_RS03360
	shikimate kinase	*aroK*	AVT04_RS01690	SMQ301_RS03410	BGL51_RS03395
	3-phosphoshikimate 1-carboxyvinyltransferase	*aroA*	AVT04_RS01695	SMQ301_RS03405	BGL51_RS03390
	chorismate synthase	*aroC*	AVT04_RS01715	SMQ301_RS03385	BGL51_RS03370
	aminodeoxychorismate/anthranilate synthase component II	*trpG*	AVT04_RS07055	SMQ301_RS07620	BGL51_RS07840
	anthranilate synthase component I	*trpE*	AVT04_RS07050	SMQ301_RS07625	BGL51_RS07845
	anthranilate phosphoribosyltransferase	*trpD*	AVT04_RS07060	SMQ301_RS07615	BGL51_RS07835
	phosphoribosylanthranilate isomerase	*trpF*	AVT04_RS07070	SMQ301_RS07605	BGL51_RS07825
	indole-3-glycerol phosphate synthase TrpC	*trpC*	AVT04_RS07065	SMQ301_RS07610	BGL51_RS07830
	tryptophan synthase subunit alpha	*trpA*	AVT04_RS07080	SMQ301_RS07595	BGL51_RS07815
	tryptophan synthase subunit beta	*trpB*	AVT04_RS07075	SMQ301_RS07600	BGL51_RS07820
Phenylalanine biosynthesis	chorismate mutase	*aroH*	AVT04_RS08945	SMQ301_RS05665	BGL51_RS05855
	prephenate dehydratase	*pheA2*	AVT04_RS01685	SMQ301_RS03415	BGL51_RS03400
	pyridoxal phosphate-dependent aminotransferase		AVT04_RS00945	SMQ301_RS04270	BGL51_RS04240
Tyrosine biosynthesis	chorismate mutase	*aroH*	AVT04_RS08945	SMQ301_RS05665	BGL51_RS05855
	prephenate dehydrogenase	*tyrA*	AVT04_RS01710	SMQ301_RS03390	BGL51_RS03375

	pyridoxal phosphate-dependent aminotransferase		AVT04_RS00945	SMQ301_RS04270	BGL51_RS04240
Alanine biosynthesis	pyridoxal phosphate-dependent aminotransferase		AVT04_RS06840	SMQ301_RS07840	BGL51_RS08065
Aspartate biosynthesis	pyridoxal phosphate-dependent aminotransferase		AVT04_RS00945	SMQ301_RS04270	BGL51_RS04240
Glutamate biosynthesis	pyridoxal phosphate-dependent aminotransferase		AVT04_RS00945	SMQ301_RS04270	BGL51_RS04240
	pyridoxal phosphate-dependent aminotransferase		AVT04_RS06840	SMQ301_RS07840	BGL51_RS08065
Lysine biosynthesis (succinyl-Dap pathway)	aspartate kinase	*lysC*	AVT04_RS02955	SMQ301_RS02095	BGL51_RS02050
	aspartate-semialdehyde dehydrogenase	*asd*	AVT04_RS08380	SMQ301_RS06300	BGL51_RS06490
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	AVT04_RS08385	SMQ301_RS06295	BGL51_RS06485
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	AVT04_RS02765	SMQ301_RS02275	BGL51_RS02220
	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase	*dapD*	AVT04_RS05890	SMQ301_RS08855	BGL51_RS08940
	acetylornithine transaminase	*argD*	AVT04_RS02560	SMQ301_RS02475	BGL51_RS02425
	succinyl-diaminopimelate desuccinylase	*dapE*	-	-	-
	diaminopimelate epimerase	*dapF*	-	-	-
	diaminopimelate decarboxylase	*lysA*	AVT04_RS03555	SMQ301_RS01465	BGL51_RS01425
	aspartate kinase	*lysC*	AVT04_RS02955	SMQ301_RS02095	BGL51_RS02050

Lysine biosynthesis (acetyl-Dap pathway)	aspartate-semialdehyde dehydrogenase	*asd*	AVT04_RS08380	SMQ301_RS06300	BGL51_RS06490
	4-hydroxy-tetrahydrodipicolinate synthase	*dapA*	AVT04_RS08385	SMQ301_RS06295	BGL51_RS06485
	4-hydroxy-tetrahydrodipicolinate reductase	*dapB*	AVT04_RS02765	SMQ301_RS02275	BGL51_RS02220
	tetrahydrodipicolinate N-acetyltransferase	*dapH*	-	-	-
	pyridoxal phosphate-dependent aminotransferase	*patA*	AVT04_RS04775	SMQ301_RS00235	BGL51_RS00210
	N-acetyldiaminopimelate deacetylase	*dapL*	AVT04_RS05895	SMQ301_RS08850	BGL51_RS08935
	diaminopimelate epimerase	*dapF*	-	-	-
	diaminopimelate decarboxylase	*lysA*	AVT04_RS03555	SMQ301_RS01465	BGL51_RS01425

Supplementary Table S7. Genes involved in urea metabolism identified in the 23 *S. thermophilus* strains

product	gene	ACA-DC 2	ASCC 1275	CNRZ1066	JIM 8232	LMD-9
acid-activated urea channel	*ureI*	STACADC2_RS01450	T303_RS02545	STR_RS01550	STH8232_RS01655	STER_RS01570
urease subunit gamma	*ureA*	STACADC2_RS01455	T303_RS02550	STR_RS01555	STH8232_RS01660	STER_RS01575
urease subunit beta	*ureB*	STACADC2_RS01460	T303_RS02555	STR_RS01560	STH8232_RS01665	STER_RS01580
urease subunit alpha	*ureC*	STACADC2_RS01465	T303_RS02560	STR_RS01565	STH8232_RS01670	STER_RS01585
urease accessory protein UreE	*ureE*	STACADC2_RS01470	T303_RS02565	STR_RS01570	STH8232_RS01675	STER_RS01590
urease accessory protein UreF	*ureF*	STACADC2_RS01475	T303_RS02570	STR_RS01575	STH8232_RS01680	STER_RS01595
urease accessory protein UreG	*ureG*	STACADC2_RS01480	T303_RS02575	STR_RS01580	STH8232_RS01685	STER_RS01600
urease accessory protein UreD	*ureD*	STACADC2_RS01485	T303_RS02580	STR_RS01585	STH8232_RS01690	STER_RS01605
cobalt transporter CbiM	*ureM*	STACADC2_RS01490	T303_RS02585	STR_RS01590	STH8232_RS01695	STER_RS01610
cobalt ABC transporter permease	*ureQ*	STACADC2_RS01495	T303_RS02590	STR_RS01595	STH8232_RS01700	STER_RS01615
ABC transporter ATP-binding protein	*ureO*	STACADC2_RS01500	T303_RS02595	STR_RS01600	STH8232_RS01705	STER_RS01620

product	gene	LMG 18311	MN-ZLW-002	ND03	APC151	B59671
acid-activated urea channel	*ureI*	STU_RS11020	Y1U_RS01490	STND_RS01520	B1761_RS07170	CG712_RS07515
urease subunit gamma	*ureA*	STU_RS11025	Y1U_RS01495	STND_RS01525	B1761_RS07175	CG712_RS07520
urease subunit beta	*ureB*	STU_RS11030	Y1U_RS01500	STND_RS01530	B1761_RS07180	CG712_RS07525
urease subunit alpha	*ureC*	STU_RS11035	Y1U_RS01505	STND_RS01535	B1761_RS07185	CG712_RS07530
urease accessory protein UreE	*ureE*	STU_RS11040	Y1U_RS01510	STND_RS01540	B1761_RS07190	CG712_RS07535
urease accessory protein UreF	*ureF*	STU_RS11045	Y1U_RS01515	STND_RS01545	B1761_RS07195	CG712_RS07540
urease accessory protein UreG	*ureG*	STU_RS11050	Y1U_RS01520	STND_RS01550	B1761_RS07200	CG712_RS07545
urease accessory protein UreD	*ureD*	STU_RS11055	Y1U_RS01525	STND_RS01555	B1761_RS07205	CG712_RS07550
cobalt transporter CbiM	*ureM*	STU_RS11060	Y1U_RS01530	STND_RS01560	B1761_RS07210	CG712_RS07555
cobalt ABC transporter permease	*ureQ*	STU_RS11065	Y1U_RS01535	STND_RS01565	B1761_RS07215	CG712_RS07560
ABC transporter ATP-binding protein	*ureO*	STU_RS11070	Y1U_RS01540	STND_RS01570	B1761_RS07220	CG712_RS07565

product	gene	CS8	DGCC 7710	EPS	GABA	KLDS 3.1003
acid-activated urea channel	*ureI*	BAY21_RS03350	CW339_RS01430	CR922_RS07915	CR921_RS01490	BEN15_RS04540
urease subunit gamma	*ureA*	BAY21_RS03345	CW339_RS01435	CR922_RS07910	CR921_RS01495	BEN15_RS04545
urease subunit beta	*ureB*	BAY21_RS03340	CW339_RS01440	CR922_RS07905	CR921_RS01500	BEN15_RS04550
urease subunit alpha	*ureC*	BAY21_RS03335	CW339_RS01445	CR922_RS07900	CR921_RS01505	BEN15_RS04555
urease accessory protein UreE	*ureE*	BAY21_RS03330	CW339_RS01450	CR922_RS07895	CR921_RS01510	BEN15_RS04560
urease accessory protein UreF	*ureF*	BAY21_RS03325	CW339_RS01455	CR922_RS07890	CR921_RS01515	BEN15_RS04565

urease accessory protein UreG	*ureG*	BAY21_RS03320	CW339_RS01460	CR922_RS07885	CR921_RS01520	BEN15_RS04570
urease accessory protein UreD	*ureD*	BAY21_RS03315	CW339_RS01465	CR922_RS07880	CR921_RS01525	BEN15_RS04575
cobalt transporter CbiM	*ureM*	BAY21_RS03310	CW339_RS01470	CR922_RS07875	CR921_RS01530	BEN15_RS04580
cobalt ABC transporter permease	*ureQ*	BAY21_RS03305	CW339_RS01475	CR922_RS07870	CR921_RS01535	BEN15_RS04585
ABC transporter ATP-binding protein	*ureO*	BAY21_RS03300	CW339_RS01480	CR922_RS07865	CR921_RS01540	BEN15_RS04590

product	**gene**	**KLDS SM**	**MN-BM-A01**	**MN-BM-A02**	**NCTC12958**	**ND07**
acid-activated urea channel	*ureI*	A9497_RS07540	AMD33_RS00445	MNA02_RS01455	DQL34_RS01805	BBD27_RS02890
urease subunit gamma	*ureA*	A9497_RS07545	AMD33_RS00450	MNA02_RS01460	DQL34_RS01810	BBD27_RS02885
urease subunit beta	*ureB*	A9497_RS07550	AMD33_RS00455	MNA02_RS01465	DQL34_RS01815	BBD27_RS02880
urease subunit alpha	*ureC*	A9497_RS07555	AMD33_RS00460	MNA02_RS01470	DQL34_RS01820	BBD27_RS02875
urease accessory protein UreE	*ureE*	A9497_RS07560	AMD33_RS00465	MNA02_RS01475	DQL34_RS01825	BBD27_RS02870
urease accessory protein UreF	*ureF*	A9497_RS07565	AMD33_RS00470	MNA02_RS01480	DQL34_RS01830	BBD27_RS02865
urease accessory protein UreG	*ureG*	A9497_RS07570	AMD33_RS00475	MNA02_RS01485	DQL34_RS01835	BBD27_RS02860
urease accessory protein UreD	*ureD*	A9497_RS07575	AMD33_RS00480	MNA02_RS01490	DQL34_RS01840	BBD27_RS02855
cobalt transporter CbiM	*ureM*	A9497_RS07580	AMD33_RS00485	MNA02_RS01495	DQL34_RS01845	BBD27_RS02850
cobalt ABC transporter permease	*ureQ*	A9497_RS07585	AMD33_RS00490	MNA02_RS01500	DQL34_RS01850	BBD27_RS02845
ABC transporter ATP-binding protein	*ureO*	A9497_RS07590	AMD33_RS00495	MNA02_RS01505	DQL34_RS01855	BBD27_RS02840

product	**gene**	**S9**	**SMQ-301**	**ST3**
acid-activated urea channel	*ureI*	AVT04_RS03430	SMQ301_RS01585	BGL51_RS01550
urease subunit gamma	*ureA*	AVT04_RS03425	SMQ301_RS01590	BGL51_RS01555
urease subunit beta	*ureB*	AVT04_RS03420	SMQ301_RS01595	BGL51_RS01560
urease subunit alpha	*ureC*	AVT04_RS03415	SMQ301_RS01600	BGL51_RS01565
urease accessory protein UreE	*ureE*	AVT04_RS03410	SMQ301_RS01605	BGL51_RS01570
urease accessory protein UreF	*ureF*	AVT04_RS03405	SMQ301_RS01610	BGL51_RS01575
urease accessory protein UreG	*ureG*	AVT04_RS03400	SMQ301_RS01615	BGL51_RS01580
urease accessory protein UreD	*ureD*	AVT04_RS03395	SMQ301_RS01620	BGL51_RS01585
cobalt transporter CbiM	*ureM*	AVT04_RS03390	SMQ301_RS01625	BGL51_RS01590
cobalt ABC transporter permease	*ureQ*	AVT04_RS03385	SMQ301_RS01630	BGL51_RS01595
ABC transporter ATP-binding protein	*ureO*	AVT04_RS03380	SMQ301_RS01635	BGL51_RS01600

Supplementary Table S8. Characteristics of confirmed Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) arrays as predicted by CRISPRFinder in the 23 *S. thermophilus* strains. Dashed lines are used to separate the two major clusters (A and B) and strain NCTC12958^T of the species, as described in the text

Strain	CRISPR-Cas systems[1]	CRISPR array coordinates	Direct Repeat (DR) consensus	DR length	Spacers
NCTC12958^T	CRISPR1	803250..805592	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	35
	CRISPR3	1568082..1569571	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	22
KLDS 3.1003	CRISPR3	121258..122019	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAACT	37	11
	CRISPR1	1254932..1255890	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	14
	CRISPR2	1506836..1507597	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	10
ASCC 1275	CRISPR1	823246..825392	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	32
	CRISPR2	1074866..1075124	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	3
	CRISPR4	1144778..1145540	GGATCACCCCCGCGTGTGCGGGAAAAAC	28	12
	CRISPR3	1558841..1559668	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	12
ND07	CRISPR1	177242..179388	GTTGTACAGTTACTTAAATCTTGAGAGTACAAAAAC	36	32
	CRISPR3	1294523..1295350	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC	36	12
	CRISPR4	1708655..1709417	GTTTTTCCCGCACACGCGGGGGTGATCC	28	12
	CRISPR2	1779078..1779330	GTCCCCTCTCGAGGTAATTAGGTTTATATC	30	3
DGCC 7710	CRISPR1	636570..638716	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	32
	CRISPR2	887984..888242	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	3
	CRISPR4	957897..958659	GGATCACCCCCGCGTGTGCGGGAAAAAC	28	12
	CRISPR3	1371965..1372792	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	12
KLDS SM	CRISPR2	199829..200087	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	3
	CRISPR4	269742..270504	GGATCACCCCCGCGTGTGCGGGAAAAAC	28	12
	CRISPR3	683810..684637	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	12
	CRISPR1	1806995..1809141	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	32
MN-BM-A02	CRISPR1	633203..635349	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	32
	CRISPR2	884614..884872	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	3
	CRISPR4	954527..955289	GGATCACCCCCGCGTGTGCGGGAAAAAC	28	12
	CRISPR3	1368586..1369413	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	12
MN-ZLW-002	CRISPR1	630165..632177	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	30
	CRISPR3	1372720..1374472	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	26
MN-BM-A01	CRISPR1	443838..445850	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	30
	CRISPR3	1186347..1188099	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	26
JIM 8232	CRISPR1	712332..715137	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	42
	CRISPR2	969817..971087	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	17
	orphan[2]	1452282..1452514	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	3
LMD-9	CRISPR1	649125..650217	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	16
	CRISPR2	897070..897328	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	3
	CRISPR3	1377229..1377794	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	8
SMQ-301	CRISPR1	649756..650847	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	16
	CRISPR2	897843..898101	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	3
	CRISPR3	1383838..1384867	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	15
ND03	CRISPR1	639512..641921	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	36
	CRISPR3	1363057..1364415	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	20
APC151	CRISPR1	1689194..1691603	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	36
	CRISPR3	575011..576303	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	19
GABA	CRISPR1	652778..655449	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	40
	CRISPR2	889434..889684	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	3
	CRISPR3	1396122..1397147	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	15
ST3	CRISPR1	652775..653929	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	17
	CRISPR2	914332..914582	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	3
	CRISPR3	1399223..1400777	GTTTTGGAACCATTCGAAACAACACAGCTCTAAAAC	36	23
LMG 18311	CRISPR1	629901..632113	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	33
	orphan[2]	774319..774515	CATATCATGCATATTGTCCATAT	23	4
	CRISPR2	863999..864323	GATATAAACCTAATTACCTCGAGAGGGGACGGAAAC	36	4
CNRZ1066	CRISPR1	625101..627844	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	41

CS8	CRISPR1	287797..290540	GTTGTACAGTTACTTAAATCTTGAGAGTACAAAAAC	36	41
S9	CRISPR1	304100..305059	GTTGTACAGTTACTTAAATCTTGAGAGTACAAAAAC	36	14
EPS	CRISPR1	1136592..1137352	GTTGTACAGTTACTTAAATCTTGAGAGTACAAAAAC	36	11
B59671	CRISPR1	1815698..1816920	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	36	18
	CRISPR4	295226..295498	GGATCACCCCCGCGTGTGCGGGAAAAAC	28	4
ACA-DC 2	-	-	-	-	-

[1]CRISPR systems are named as described by Horvath and Barrangou, 2010
[2]CRISPR array identified in the absence of Cas proteins

Supplementary Table S9A. Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated (CRISPR-Cas) system types and CRISPR-Cas genes identified in the 23 *S. thermophilus* strains. Genes highlighted in red correspond to putative pseudogenes while genes highlighted in blue are truncated

CRISPR-Cas system type		ACA-DC 2	APC151	ASCC 1275	B59671
CRISPR1 (Class 2 / II-A)	**coordinates of CRISPR array**	-	**1689194.. 1691603**	**823246..825392**	**1815698..1816920**
	type II CRISPR RNA-guided endonuclease Cas9	STACADC2_RS03320	B1761_RS09095	T303_RS04490	CG712_RS09600
	subtype II CRISPR-associated endonuclease Cas1	STACADC2_RS03325	B1761_RS09100	T303_RS04495	CG712_RS09605
	CRISPR-associated endonuclease Cas2	STACADC2_RS03330	B1761_RS09105	T303_RS04500	CG712_RS09610
	CRISPR-associated protein Csn2	STACADC2_RS03335	B1761_RS09110	T303_RS04505	CG712_RS09615
CRISPR2 (Class 1 / III-A)*	**coordinates of CRISPR array**	**849603..849704**	**85936..86037**	**1074866..1075124**	-
	CRISPR-associated endonuclease Cas1	STACADC2_RS04550	B1761_RS00500	T303_RS05790	CG712_RS01090
	CRISPR-associated endonuclease Cas2	STACADC2_RS04555	B1761_RS00505	T303_RS05795	CG712_RS01095
	CRISPR-associated endoribonuclease Cas6	STACADC2_RS04560	B1761_RS00510	T303_RS05800	CG712_RS01100
	type III-A CRISPR-associated protein Cas10/Csm1	STACADC2_RS04565	B1761_RS00515	T303_RS05805	CG712_RS01105
	type III-A CRISPR-associated protein Csm2	-	B1761_RS00520	T303_RS05810	CG712_RS01110
	type III-A CRISPR-associated protein Csm2	STACADC2_RS04570	-	-	-
	type III-A CRISPR-associated RAMP protein Csm3	STACADC2_RS04575	B1761_RS00525	T303_RS05815	CG712_RS01115
	type III-A CRISPR-associated RAMP protein Csm4	STACADC2_RS04580	B1761_RS00530	T303_RS05820	CG712_RS01120
	type III-A CRISPR-associated RAMP protein Csm5	STACADC2_RS04585	B1761_RS00535	T303_RS05825	CG712_RS01125
	type III-A CRISPR-associated protein Csm6	STACADC2_RS04590	-	-	-
	hypothetical protein	-	-	-	-
	type III-A CRISPR-associated protein Csm6	-	-	-	-
	type III-A CRISPR-associated protein Csm6	-	B1761_RS00540	T303_RS05830	CG712_RS01130
CRISPR3 (Class 2 / II-A)	**coordinates of CRISPR array**	-	**575011.. 576303**	**1558841..1559668**	-
	type II-A CRISPR-associated protein Csn2	-	B1761_RS03065	T303_RS08255	-
	CRISPR-associated endonuclease Cas2	-	B1761_RS03070	T303_RS08260	-
	subtype II CRISPR-associated endonuclease Cas1	-	B1761_RS03075	T303_RS08265	-
	type II CRISPR RNA-guided endonuclease Cas9	-	B1761_RS03080	T303_RS08270	-
CRISPR4 (Class 1 / I-E)	**coordinates of CRISPR array**	-	-	**1144778..1145540**	**295226..295498**
	type I-E CRISPR-associated endoribonuclease Cas2	-	-	T303_RS06145	CG712_RS01545
	subtype I-E CRISPR-associated endonuclease Cas1	-	-	T303_RS06150	CG712_RS01550
	type I-E CRISPR-associated protein Cas6/Cse3/CasE	-	-	T303_RS06155	CG712_RS01555
	type I-E CRISPR-associated protein Cas5/CasD	-	-	T303_RS06160	CG712_RS01560

	type I-E CRISPR-associated protein Cas7/Cse4/CasC	-	-	T303_RS06165	CG712_RS01565
	type I-E CRISPR-associated protein Cse2/CasB	-	-	T303_RS06170	CG712_RS01570
	CRISPR-associated protein Cse1/Cas8e/CasA	-	-	T303_RS06175	CG712_RS01575
	CRISPR-associated helicase/endonuclease Cas3	-	-	T303_RS06180	CG712_RS01580

CRISPR-Cas system type		**CNRZ1066**	**CS8**	**DGCC 7710**	**EPS**
CRISPR1 (Class 2 / II-A)	**coordinates of CRISPR array**	**625101..627844**	**287797..290540**	**636570..638716**	**1136592..1137352**
	type II CRISPR RNA-guided endonuclease Cas9	STR_RS03365	BAY21_RS01550	CW339_RS03425	CR922_RS06085
	subtype II CRISPR-associated endonuclease Cas1	STR_RS03370	BAY21_RS01545	CW339_RS03430	CR922_RS06080
	CRISPR-associated endonuclease Cas2	STR_RS03375	BAY21_RS01540	CW339_RS03435	CR922_RS06075
	CRISPR-associated protein Csn2	STR_RS03380	BAY21_RS01535	CW339_RS03440	CR922_RS06070
CRISPR2 (Class 1 / III-A)*	**coordinates of CRISPR array**	-	-	**887984..888242**	**896474..896575**
	CRISPR-associated endonuclease Cas1	STR_RS04695	BAY21_RS00240	CW339_RS04755	CR922_RS04745
	CRISPR-associated endonuclease Cas2	STR_RS04705	BAY21_RS00230	CW339_RS04760	CR922_RS04740
	CRISPR-associated endoribonuclease Cas6	-	-	CW339_RS04765	CR922_RS04735
	type III-A CRISPR-associated protein Cas10/Csm1	-	-	CW339_RS04770	CR922_RS04730
	type III-A CRISPR-associated protein Csm2	STR_RS09985	BAY21_RS09445	CW339_RS04775	CR922_RS04725
	type III-A CRISPR-associated protein Csm2	-	-	-	-
	type III-A CRISPR-associated RAMP protein Csm3	STR_RS04710	BAY21_RS00225	CW339_RS04780	CR922_RS04720
	type III-A CRISPR-associated RAMP protein Csm4	STR_RS04715	BAY21_RS00220	CW339_RS04785	CR922_RS04715
	type III-A CRISPR-associated RAMP protein Csm5	STR_RS04720	BAY21_RS00215	CW339_RS04790	CR922_RS04710
	type III-A CRISPR-associated protein Csm6	-	-	-	-
	hypothetical protein	-	-	-	-
	type III-A CRISPR-associated protein Csm6	-	-	-	-
	type III-A CRISPR-associated protein Csm6	STR_RS04725	BAY21_RS00210	CW339_RS04795	CR922_RS04705
CRISPR3 (Class 2 / II-A)	**coordinates of CRISPR array**	-	-	**1371965..1372792**	-
	type II-A CRISPR-associated protein Csn2	-	-	CW339_RS07265	-
	CRISPR-associated endonuclease Cas2	-	-	CW339_RS07270	-
	subtype II CRISPR-associated endonuclease Cas1	-	-	CW339_RS07275	-
	type II CRISPR RNA-guided endonuclease Cas9	-	-	CW339_RS07280	-
CRISPR4 (Class 1 / I-E)	**coordinates of CRISPR array**	-	-	**957897..958659**	-
	type I-E CRISPR-associated endoribonuclease Cas2	-	-	CW339_RS05115	-
	subtype I-E CRISPR-associated endonuclease Cas1	-	-	CW339_RS05120	-

	type I-E CRISPR-associated protein Cas6/Cse3/CasE	-	-	CW339_RS05125	-
	type I-E CRISPR-associated protein Cas5/CasD	-	-	CW339_RS05130	-
	type I-E CRISPR-associated protein Cas7/Cse4/CasC	-	-	CW339_RS05135	-
	type I-E CRISPR-associated protein Cse2/CasB	-	-	CW339_RS05140	-
	CRISPR-associated protein Cse1/Cas8e/CasA	-	-	CW339_RS05145	-
	CRISPR-associated helicase/endonuclease Cas3	-	-	CW339_RS05150	-

CRISPR-Cas system type		**GABA**	**JIM 8232**	**KLDS 3.1003**	**KLDS SM**
CRISPR1 (Class 2 / II-A)	**coordinates of CRISPR array**	**652778..655449**	**712332..715137**	**1254932..1255890**	**1806995..1809141**
	type II CRISPR RNA-guided endonuclease Cas9	CR921_RS03475	STH8232_RS03750	BEN15_RS06675	A9497_RS09510
	subtype II CRISPR-associated endonuclease Cas1	CR921_RS03480	STH8232_RS03755	BEN15_RS06680	A9497_RS09515
	CRISPR-associated endonuclease Cas2	CR921_RS03485	STH8232_RS03760	BEN15_RS06685	A9497_RS09520
	CRISPR-associated protein Csn2	CR921_RS03490	STH8232_RS03765	BEN15_RS06690	A9497_RS09525
CRISPR2 (Class 1 / III-A)*	**coordinates of CRISPR array**	**889434..889684**	**969817..971087**	**1506836..1507597**	**199829..200087**
	CRISPR-associated endonuclease Cas1	CR921_RS04685	STH8232_RS05055	BEN15_RS07965	A9497_RS01025
	CRISPR-associated endonuclease Cas2	CR921_RS04690	STH8232_RS05060	BEN15_RS07970	A9497_RS01030
	CRISPR-associated endoribonuclease Cas6	-	STH8232_RS05065	BEN15_RS07975	A9497_RS01035
	type III-A CRISPR-associated protein Cas10/Csm1	-	STH8232_RS05070	BEN15_RS07980	A9497_RS01040
	type III-A CRISPR-associated protein Csm2	-	-	-	A9497_RS01045
	type III-A CRISPR-associated protein Csm2	-	STH8232_RS05075	BEN15_RS07985	-
	type III-A CRISPR-associated RAMP protein Csm3	-	STH8232_RS05080	BEN15_RS07990	A9497_RS01050
	type III-A CRISPR-associated RAMP protein Csm4	-	STH8232_RS05085	BEN15_RS07995	A9497_RS01055
	type III-A CRISPR-associated RAMP protein Csm5	-	STH8232_RS05090	BEN15_RS08000	A9497_RS01060
	type III-A CRISPR-associated protein Csm6	CR921_RS04825	STH8232_RS05095	-	-
	hypothetical protein	-	-	-	-
	type III-A CRISPR-associated protein Csm6	CR921_RS04830	STH8232_RS05100	-	-
	type III-A CRISPR-associated protein Csm6	-	-	BEN15_RS08005	A9497_RS01065
CRISPR3 (Class 2 / II-A)	**coordinates of CRISPR array**	**1396122..1397147**	**-**	**121258..122019**	**683810..684637**
	type II-A CRISPR-associated protein Csn2	CR921_RS07345	-	BEN15_RS00605	A9497_RS03470
	CRISPR-associated endonuclease Cas2	CR921_RS07350	-	BEN15_RS00610	A9497_RS03475
	subtype II CRISPR-associated endonuclease Cas1	CR921_RS07355	-	BEN15_RS00615	A9497_RS03480
	type II CRISPR RNA-guided endonuclease Cas9	CR921_RS07360	-	BEN15_RS00620	A9497_RS03485
CRISPR4 (Class 1 / I-E)	**coordinates of CRISPR array**	**-**	**-**	**-**	**269742..270504**

	type I-E CRISPR-associated endoribonuclease Cas2	-	-	-	A9497_RS01385
	subtype I-E CRISPR-associated endonuclease Cas1	-	-	-	A9497_RS01390
	type I-E CRISPR-associated protein Cas6/Cse3/CasE	-	-	-	A9497_RS01395
	type I-E CRISPR-associated protein Cas5/CasD	-	-	-	A9497_RS01400
	type I-E CRISPR-associated protein Cas7/Cse4/CasC	-	-	-	A9497_RS01405
	type I-E CRISPR-associated protein Cse2/CasB	-	-	-	A9497_RS01410
	CRISPR-associated protein Cse1/Cas8e/CasA	-	-	-	A9497_RS01415
	CRISPR-associated helicase/endonuclease Cas3	-	STH8232_RS05405	-	A9497_RS01420

CRISPR-Cas system type		**LMD-9**	**LMG 18311**	**MN-BM-A01**	**MN-BM-A02**
CRISPR1 (Class 2 / II-A)	**coordinates of CRISPR array**	**649125..650217**	**629901..632113**	**443838..445850**	**633203..635349**
	type II CRISPR RNA-guided endonuclease Cas9	STER_RS03490	STU_RS12835	AMD33_RS02300	MNA02_RS03410
	subtype II CRISPR-associated endonuclease Cas1	STER_RS03495	STU_RS12840	AMD33_RS02305	MNA02_RS03415
	CRISPR-associated endonuclease Cas2	STER_RS03500	STU_RS12845	AMD33_RS02310	MNA02_RS03420
	CRISPR-associated protein Csn2	STER_RS03505	STU_RS12850	AMD33_RS02315	MNA02_RS03425
CRISPR2 (Class 1 / III-A)*	**coordinates of CRISPR array**	**897070..897328**	**863999..864323**	**719325..719426**	**884614..884872**
	CRISPR-associated endonuclease Cas1	STER_RS04790	STU_RS14035	AMD33_RS03795	MNA02_RS04700
	CRISPR-associated endonuclease Cas2	STER_RS04795	STU_RS14045	AMD33_RS03790	MNA02_RS04705
	CRISPR-associated endoribonuclease Cas6	STER_RS04800	STU_RS14050	AMD33_RS03785	MNA02_RS04710
	type III-A CRISPR-associated protein Cas10/Csm1	STER_RS04805	STU_RS14055	AMD33_RS03780	MNA02_RS04715
	type III-A CRISPR-associated protein Csm2	STER_RS04810	STU_RS14060	AMD33_RS03775	MNA02_RS04720
	type III-A CRISPR-associated protein Csm2	-	-	-	-
	type III-A CRISPR-associated RAMP protein Csm3	STER_RS04815	STU_RS14065	AMD33_RS03770	MNA02_RS04725
	type III-A CRISPR-associated RAMP protein Csm4	STER_RS04820	STU_RS14070	AMD33_RS03765	MNA02_RS04730
	type III-A CRISPR-associated RAMP protein Csm5	STER_RS04825	STU_RS14075	AMD33_RS03760	MNA02_RS04735
	type III-A CRISPR-associated protein Csm6	-	-	-	-
	hypothetical protein	-	STU_RS14080	-	-
	type III-A CRISPR-associated protein Csm6	-	-	-	-
	type III-A CRISPR-associated protein Csm6	STER_RS04830	STU_RS14085	AMD33_RS03755	MNA02_RS04740
CRISPR3 (Class 2 / II-A)	**coordinates of CRISPR array**	**1377229..1377794**	**-**	**1186347..1188099**	**1368586..1369413**
	type II-A CRISPR-associated protein Csn2	STER_RS07245	-	AMD33_RS06155	MNA02_RS07160
	CRISPR-associated endonuclease Cas2	STER_RS07250	-	AMD33_RS06160	MNA02_RS07165
	subtype II CRISPR-associated endonuclease Cas1	STER_RS07255	-	AMD33_RS06165	MNA02_RS07170

	type II CRISPR RNA-guided endonuclease Cas9	STER_RS07260	-	AMD33_RS06170	MNA02_RS07175
CRISPR4 (Class 1 / I-E)	**coordinates of CRISPR array**	**-**	**-**	**-**	**954527..955289**
	type I-E CRISPR-associated endoribonuclease Cas2	-	-	-	MNA02_RS05055
	subtype I-E CRISPR-associated endonuclease Cas1	-	-	-	MNA02_RS05060
	type I-E CRISPR-associated protein Cas6/Cse3/CasE	-	-	-	MNA02_RS05065
	type I-E CRISPR-associated protein Cas5/CasD	-	-	-	MNA02_RS05070
	type I-E CRISPR-associated protein Cas7/Cse4/CasC	-	-	-	MNA02_RS05075
	type I-E CRISPR-associated protein Cse2/CasB	-	-	-	MNA02_RS05080
	CRISPR-associated protein Cse1/Cas8e/CasA	-	-	-	MNA02_RS05085
	CRISPR-associated helicase/endonuclease Cas3	-	-	-	MNA02_RS05090

CRISPR-Cas system type		**MN-ZLW-002**	**NCTC12958**	**ND03**	**ND07**
CRISPR1 (Class 2 / II-A)	**coordinates of CRISPR array**	**630165..632177**	**803250..805592**	**639512..641921**	**177242..179388**
	type II CRISPR RNA-guided endonuclease Cas9	Y1U_RS03345	DQL34_RS04340	STND_RS03420	BBD27_RS00925
	subtype II CRISPR-associated endonuclease Cas1	Y1U_RS03350	DQL34_RS04345	STND_RS03425	BBD27_RS00920
	CRISPR-associated endonuclease Cas2	Y1U_RS03355	DQL34_RS04350	STND_RS03430	BBD27_RS00915
	CRISPR-associated protein Csn2	Y1U_RS03360	DQL34_RS04355	STND_RS03435	BBD27_RS00910
CRISPR2 (Class 1 / III-A)*	**coordinates of CRISPR array**	**905670..905771**	**-**	**873997..874098**	**1779078..1779330**
	CRISPR-associated endonuclease Cas1	Y1U_RS04810	-	STND_RS04630	BBD27_RS09325
	CRISPR-associated endonuclease Cas2	Y1U_RS04805	-	STND_RS04635	BBD27_RS09320
	CRISPR-associated endoribonuclease Cas6	Y1U_RS04800	-	STND_RS04640	BBD27_RS09315
	type III-A CRISPR-associated protein Cas10/Csm1	Y1U_RS04795	-	STND_RS04645	BBD27_RS09310
	type III-A CRISPR-associated protein Csm2	Y1U_RS04790	-	STND_RS04650	BBD27_RS09305
	type III-A CRISPR-associated protein Csm2	-	-	-	-
	type III-A CRISPR-associated RAMP protein Csm3	Y1U_RS04785	-	STND_RS04655	BBD27_RS09300
	type III-A CRISPR-associated RAMP protein Csm4	Y1U_RS04780	-	STND_RS04660	BBD27_RS09295
	type III-A CRISPR-associated RAMP protein Csm5	Y1U_RS04775	-	STND_RS04665	BBD27_RS09290
	type III-A CRISPR-associated protein Csm6	-	-	-	-
	hypothetical protein	-	-	-	-
	type III-A CRISPR-associated protein Csm6	-	-	-	-
	type III-A CRISPR-associated protein Csm6	Y1U_RS04770	-	STND_RS04670	BBD27_RS09285
CRISPR3 (Class 2 / II-A)	**coordinates of CRISPR array**	**1372720..1374472**	**1568082..1569571**	**1363057..1364415**	**1294523..1295350**
	type II-A CRISPR-associated protein Csn2	Y1U_RS07175	DQL34_RS08310	STND_RS07165	BBD27_RS06880

	CRISPR-associated endonuclease Cas2	Y1U_RS07180	DQL34_RS08315	STND_RS07170	BBD27_RS06875
	subtype II CRISPR-associated endonuclease Cas1	Y1U_RS07185	DQL34_RS08320	STND_RS07175	BBD27_RS06870
	type II CRISPR RNA-guided endonuclease Cas9	Y1U_RS07190	DQL34_RS08325	STND_RS07180	BBD27_RS06865
CRISPR4 (Class 1 / I-E)	**coordinates of CRISPR array**	-	-	-	**1708655..1709417**
	type I-E CRISPR-associated endoribonuclease Cas2	-	-	-	BBD27_RS08965
	subtype I-E CRISPR-associated endonuclease Cas1	-	-	-	BBD27_RS08960
	type I-E CRISPR-associated protein Cas6/Cse3/CasE	-	-	-	BBD27_RS08955
	type I-E CRISPR-associated protein Cas5/CasD	-	-	-	BBD27_RS08950
	type I-E CRISPR-associated protein Cas7/Cse4/CasC	-	-	-	BBD27_RS08945
	type I-E CRISPR-associated protein Cse2/CasB	-	-	-	BBD27_RS08940
	CRISPR-associated protein Cse1/Cas8e/CasA	-	-	-	BBD27_RS08935
	CRISPR-associated helicase/endonuclease Cas3	-	-	-	BBD27_RS08930

CRISPR-Cas system type		**S9**	**SMQ-301**	**ST3**
CRISPR1 (Class 2 / II-A)	**coordinates of CRISPR array**	**304100..305059**	**649756..650847**	**652775..653929**
	type II CRISPR RNA-guided endonuclease Cas9	AVT04_RS01630	SMQ301_RS03470	BGL51_RS03460
	subtype II CRISPR-associated endonuclease Cas1	AVT04_RS01625	SMQ301_RS03475	BGL51_RS03465
	CRISPR-associated endonuclease Cas2	AVT04_RS01620	SMQ301_RS03480	BGL51_RS03470
	CRISPR-associated protein Csn2	AVT04_RS01615	SMQ301_RS03485	BGL51_RS03475
CRISPR2 (Class 1 / III-A)*	**coordinates of CRISPR array**	**70568..70669**	**897843..898101**	**914332..914582**
	CRISPR-associated endonuclease Cas1	AVT04_RS00360	SMQ301_RS04770	BGL51_RS04890
	CRISPR-associated endonuclease Cas2	AVT04_RS00350	SMQ301_RS04775	BGL51_RS04895
	CRISPR-associated endoribonuclease Cas6	AVT04_RS00345	SMQ301_RS04780	-
	type III-A CRISPR-associated protein Cas10/Csm1	AVT04_RS00340	SMQ301_RS10100	-
	type III-A CRISPR-associated protein Csm2	AVT04_RS00335	SMQ301_RS04790	-
	type III-A CRISPR-associated protein Csm2	-	-	-
	type III-A CRISPR-associated RAMP protein Csm3	AVT04_RS00330	SMQ301_RS04795	-
	type III-A CRISPR-associated RAMP protein Csm4	AVT04_RS00325	SMQ301_RS04800	-
	type III-A CRISPR-associated RAMP protein Csm5	AVT04_RS00320	SMQ301_RS04805	-
	type III-A CRISPR-associated protein Csm6	-	-	BGL51_RS05025
	hypothetical protein	-	-	-
	type III-A CRISPR-associated protein Csm6	-	-	BGL51_RS05030
	type III-A CRISPR-associated protein Csm6	AVT04_RS00315	SMQ301_RS10105	-

CRISPR3 (Class 2 / II-A)	**coordinates of CRISPR array**	**-**	**1383838..1384867**	**1399223..1400777**
	type II-A CRISPR-associated protein Csn2	-	SMQ301_RS07230	BGL51_RS07450
	CRISPR-associated endonuclease Cas2	-	SMQ301_RS07235	BGL51_RS07455
	subtype II CRISPR-associated endonuclease Cas1	-	SMQ301_RS07240	BGL51_RS07460
	type II CRISPR RNA-guided endonuclease Cas9	-	SMQ301_RS07245	BGL51_RS07465
CRISPR4 (Class 1 / I-E)	**coordinates of CRISPR array**	**-**	**-**	**-**
	type I-E CRISPR-associated endoribonuclease Cas2	-	-	-
	subtype I-E CRISPR-associated endonuclease Cas1	-	-	-
	type I-E CRISPR-associated protein Cas6/Cse3/CasE	-	-	-
	type I-E CRISPR-associated protein Cas5/CasD	-	-	-
	type I-E CRISPR-associated protein Cas7/Cse4/CasC	-	-	-
	type I-E CRISPR-associated protein Cse2/CasB	-	-	-
	CRISPR-associated protein Cse1/Cas8e/CasA	-	-	-
	CRISPR-associated helicase/endonuclease Cas3	-	SMQ301_RS05125	-

*****CRISPR2 systems highlighted in green correspond to possible CRISPRs as identified by CRISPRFinder and contain only 1 spacer between two repeats

Supplementary Table S9B. Blastn analysis of the spacers of the various Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated (CRISPR-Cas) systems predicted in the 23 *S. thermophilus* strains

Spacers of CRISPR1		
Strain (coordinates of CRISPR array)	**spacers**	**1st blastn result (excluding same species)**
LMG 18311 (629901..632113)	>spacer1	
	GAGCTACCAGCTACCCCGTATGTCAGAGAG	*Streptococcus thermophilus* bacteriophage 7201
	>spacer2	
	CGTTCCTTTTTTCAAGGTAATCTTTGAAAG	
	>spacer3	
	AAGTCCGTAAGCACCAGTTCCAATCGTCAT	*Streptococcus* phage D4276
	>spacer4	
	TTGAATACCAATGCCAGCTTCTTTTAAGGC	*Streptococcus* phages: 9A / P9854 / P9851 / P8922 / P8921 / P7574 / P7154 / P7152 / P7151 / P7134 / P7133 / P5652 / P5651 / P3681 / P0094 / P0093 / P0092 / CHPC1151 / CHPC926 / 5093
	>spacer5	
	AACCTCATACATGGGGAAAATTGGTAAGTA	*Streptococcus* phage D4276
	>spacer6	
	TAACTTCATTAGTGTAGTTGTAATTAGCAT	
	>spacer7	
	TTAGCTACCCAAATATCTTCTGTTTTCCAA	*Streptococcus* phage D4276
	>spacer8	
	GAGTTTTCAATATTGGCACAGGAGACAATT	
	>spacer9	
	TGATACTATTTTAGTCAGATATGAAATATC	
	>spacer10	
	TCATCAATGTTTAAAGCCCAACAATACATGA	
	>spacer11	
	TAGATTTAATCAGTAATGAGTTAGGCATAA	*Streptococcus* phages: M19 / vB_SthS_VA214 / P4761 / TP-778L / Abc2
	>spacer12	
	AGGAAAATAGCATGAGCGTACAACAATCTA	*Streptococcus* phage P5652
	>spacer13	

	TGTCTATCACGCTTCCTAAGTGCATGAAAA	*Streptococcus thermophilus* bacteriophages: Sfi11 / 7201 / Sfi21 / Sfi19
	>spacer14	
	ATGTCACCAATCACTAAAGAACCTACGCTG	*Streptococcus* phages: phiARI0468b-3 / TP-778L / TP-J34; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer15	
	AACATCTTCCTCTCCGATTGCAAATAGTGC	*Streptococcus* phages: STP2 / STP1 / L5A1 / C0 / B5 / B0 / 7A5 / A0 / vB_SthS_VA214 / P4761
	>spacer16	
	CATATTTGGTGCCCGTTCGATAAAGAGTA	*Streptococcus thermophilus* bacteriophage Sfi21
	>spacer17	
	CATTAAATCGCTTGAAGCAGACATTGAAGC	*Streptococcus* phage MM25
	>spacer18	
	GACTTATCTTGGAAGGTAGTGAAGGCACTT	
	>spacer19	
	TCCTTGCCATCTGCACTGTAAGCCCAAGCA	*Streptococcus thermophilus* bacteriophage Sfi11
	>spacer20	
	TAGTACGCATAATCAATTCATCAAGCTTGA	
	>spacer21	
	GTAGTGACCCAAAATTCTATGACCTTGAAA	*Streptococcus thermophilus* bacteriophage Sfi21
	>spacer22	
	AGATTGTGGTGCTTACGGAAAATTCCTTGT	*Streptococcus* phage P9903
	>spacer23	
	TGGCAAGAAGTGTAAGAGATGCAATGGATA	*Streptococcus thermophilus* bacteriophage O1205
	>spacer24	
	TTTATTATCATTATTCTTCTTCCCAAGCGT	
	>spacer25	
	TTTTATAGAATTTGGTGGTGAACTTTTTCA	
	>spacer26	
	AATGGGTCACAGATTGCCATAATAAGGAG	
	>spacer27	

	CCGAGGTCACTTTAGAACCCACAAAATAAG	*Streptococcus thermophilus* bacteriophage Sfi19
	>spacer28	
	ATGAGAGAACACAGTATAGACCCTGATACA	*Streptococcus* phage vB_SthS_VA214; *Streptococcus thermophilus* bacteriophages: 7201 / Sfi19
	>spacer29	
	CAGTATTAATGAGGTTTGGGTGGTCATTCC	*Streptococcus* phages: P7602 / P7601 / P7572 / P7571 / P5641 / P4761 / 73
	>spacer30	
	CCATACTCTCTATCAGTTCATTTAATTCTTC	
	>spacer31	
	TAATATGTCGCTCTACTGATTCCAAAACGG	
	>spacer32	
	ATGAATTACATTCATGATTTTATCGAGTTT	
	>spacer33	
	CGTGCCATTGTTTCGGTCGGACGTGGGCA	*Streptococcus* phage P5652
EPS (1136592..1137352) (rev_com)	>spacer1	
	CGTGCCATTGTTTCGGTCGGACGTGGGCA	*Streptococcus* phage P5652
	>spacer2	
	ATGAGAGAACACAGTATAGACCCTGATACA	*Streptococcus* phage vB_SthS_VA214; *Streptococcus thermophilus* bacteriophage 7201 / Sfi19
	>spacer3	
	CCGAGGTCACTTTAGAACCCACAAAATAAG	*Streptococcus thermophilus* bacteriophage Sfi19
	>spacer4	
	AATGGGTCACAGATTGCCATAATAAGGAGG	
	>spacer5	
	TTTTATAGAATTTGGTGGTGAACTTTTTCA	
	>spacer6	
	TTTATTATCATTATTCTTCTTCCCAAGCGT	
	>spacer7	
	TGGCAAGAAGTGTAAGAGATGCAATGGATA	*Streptococcus thermophilus* bacteriophage O1205
	>spacer8	
	AGATTGTGGTGCTTACGGAAAATTCCTTGT	*Streptococcus* phage P9903

	>spacer9	
	GTAGTGACCCAAAATTCTATGACCTTGAAA	*Streptococcus thermophilus* bacteriophage Sfi21
	>spacer10	
	TAGTACGCATAATCAATTCATCAAGCTTGA	
	>spacer11	
	TAATTCTTTCAGGTATGAAACTAGAAACAC	
S9 (304100..305059) (rev_com)	>spacer1	
	CTTTCTAAGTTGAATTAAATTCAAGTTTTG	*Streptococcus* phages: TP-778L / TP-J34
	>spacer2	
	TCGCTACTATGGTTAACGATGAGGAACTCT	*Streptococcus* phages: P4761 / 73 / TP-778L / TP-J34
	>spacer3	
	AGCAACTTTAAAACTAAAAGAGCTACTTGA	*Streptococcus thermophilus* bacteriophage Sfi19
	>spacer4	
	CAGTATTAATGAGGTTTGGGTGGTCATTCC	*Streptococcus* phages: P7602 / P7601 / P7572 / P7571 / P5641 / P4761 / 73
	>spacer5	
	CCATACTCTCTATCAGTTCATTTAATTCTTC	
	>spacer6	
	TGAGAGTGTCTGATGGATTTATTGGTAACC	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer7	
	ACCTCATACATGGGGAAAACTTGTAAGTA	*Streptococcus* phage D4276
	>spacer8	
	TATTTCACGAATTTCTACACTTTTCAACCT	
	>spacer9	
	CTGAAACCTTGTTTTGAAGCGCTTGGAAGT	*Streptococcus* phage TP-J34
	>spacer10	
	GTCAATTGATACTGCAATCTCTTTAACATT	
	>spacer11	
	ACTTCAATATGGTCAACATCTTGATCACCGA	
	>spacer12	
	TAATATGTCGCTCTACTGATTCCAAAACGG	
	>spacer13	
	ATGAATTACATTCATGATTTTATCGAGTTT	

	>spacer14	
	CGTGCCATTGTTTCGGTCGGACGTGGGCA	*Streptococcus* phage P5652
CNRZ1066 (625101..627844)	>spacer1	
	AGAACGTATTCCAAAACCTCTTTACGATTA	*Streptococcus* phages: V2 / R1 / MM25 / 31B4 / 16B8 / 9B4 / 9A / 7T / P9854 / P9853 / P9852 / P9851 / P8922 / P8921 / P7954 / P7953 / P7952 / P7633 / P7632 / P7631 / P7602 / P7601 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P3684 / P3681 / 128 / 53 / CHPC577 / ALQ13.2 / vB_SthS_VA460 / vB_SthS_VA698; *Streptococcus thermophilus* bacteriophages: kappa3 / 2972 / DT1
	>spacer2	
	TTAACTGTTATCAAAATGATAAGATAGTCT	
	>spacer3	
	CGTTGATGTTTATTCAAGTAAAATAATTAA	*Streptococcus thermophilus* temperate bacteriophage O1205
	>spacer4	
	TCCTTTCACGGGTAGCACACTAACATACAC	
	>spacer5	
	GTTGGCAATGCAAACAACCTTTATGAACCG	*Streptococcus* phage TP-J34; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer6	
	TTTATTTCCTTGCGATAACGTTCCACCTTT	*Streptococcus* phage TP-J34; *Streptococcus thermophilus* bacteriophages: Sfi11 / O1205
	>spacer7	
	AGATTATAAGGAACACAACCAACTATATAG	
	>spacer8	
	ACGACATCAAGCTGATTGTCTTCTACATAA	*Streptococcus* phages: STP1 / B5 / P9902 / 73; *Streptococcus thermophilus* bacteriophage DT1
	>spacer9	
	TTTGGAATACTGAATGTTTTACTGAAAATC	
	>spacer10	
	ACACCACTATCTTTTCCTCCTGAAAATGAA	
	>spacer11	

	GTAATTCCACGAAATTATCAACCTTATGCA	*Streptococcus* phage P9903 / P5641 / P4761 / D4276 / 73; *Streptococcus thermophilus* bacteriophages: DT1 / Sfi21
	>spacer12	
	TTGGAGGATTGCCCCATATTCCCAAGAGT	
	>spacer13	
	GAGAGGCGTTAAATATAGAAATGCAAGATT	*Streptococcus* phage vB_SthS_VA698; *Streptococcus thermophilus* temperate bacteriophage O1205
	>spacer14	
	TTTTAACGTCATCAGTCCACCGCCTTAAAT	
	>spacer15	
	CACCTCTTTCGATGGAAAGGTATCCTTCTA	*Streptococcus* phage vB_SthS_VA460
	>spacer16	
	GACCAAAGTTTGATTATAGAGCTATACACC	
	>spacer17	
	ACCATCATTCTTACCATTACAACTGTAATG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer18	
	ATACGAATTCGGTTCGCACAATTACAATTC	
	>spacer19	
	TATCAACGCAATCATTACAACAACTTCAAACA	
	>spacer20	
	ATCTACGTGTCAATACATATCACAAAACAG	*Streptococcus* phages: P7954 / pC2S4; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer21	
	ATTTTTAGAAATTTCTGATATAATAATGA	
	>spacer22	
	TTGTTGGAACAAGGACGACTTGGTAAACTA	*Streptococcus* virus 9874
	>spacer23	
	CATATTAAGCTGACTGGGCCTAATGCTTTT	*Streptococcus* phages: P9854 / P9852 / P9851 / 858 / TP-778L; *Lactococcus* phages: AM5 / AM4; *Streptococcus thermophilus* bacteriophages: 2972 / O1205
	>spacer24	
	TTCATAGCATACCGTAGTTGTAAAATCTAT	
	>spacer25	

	AACATTTAGGGAATGAAATTGATAAGACTG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer26	
	AACATGAGAAACTGTAGAAAACAAGCAATA	
	>spacer27	
	TGGTGAAGATGGCAGTCATAAATGGCACATT	*Streptococcus* phages: vB_SthS_VA214 / P9851 / P0091
	>spacer28	
	AAGGGTTGAAAAATGTTGGTATATCAAACG	*Streptococcus* phages: V2 / R1 / L5A1 / C0 / 31B4 / 9B4 / A0 / P7633 / P7632 / P7631 / P7574
	>spacer29	
	TTCTGGTAGTGGATTTAGTCAAACAGATGT	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer30	
	TCCATAGAGCGTCTTAAACAAAGAATAGTC	
	>spacer31	
	GTTGGCAATGCAAACAACCTTTATGAACCG	*Streptococcus* phage TP-J34; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer32	
	CATATTAAGCTGACTGGGCCTAATGCTTTT	*Streptococcus* phages: P9854 / P9852 / P9851 / 858 / TP-778L; *Lactococcus* phages: AM5 / AM4; *Streptococcus thermophilus* bacteriophages: 2972 / O1205
	>spacer33	
	TTCATAGCATACCGTAGTTGTAAAATCTAT	
	>spacer34	
	AACATTTAGGGAATGAAATTGATAAGACTG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer35	
	AACATGAGAAACTGTAGAAAACAAGCAATA	
	>spacer36	
	TGGTGAAGATGGCAGTCATAAATGGCACATT	*Streptococcus* phages: vB_SthS_VA214 / P9851 / P0091
	>spacer37	
	TTATGATTGAATGACATGGTTGTATAAGTA	*Streptococcus* phages: P9854 / P9852 / P9851 / TP-778L / 858; *Streptococcus thermophilus* bacteriophage 2972

	>spacer38	
	TTTCTTTAGGAATACCAGGGAGTTCAGCTT	*Enterococcus faecium* HB6 plasmid pAR6 / *Enterococcus faecium* JH95 plasmid pJS33 / *Enterococcus faecium* plasmid pEFR
	>spacer39	
	TGGCAGAGATTACACAGCAACGGAAACAGC	*Streptococcus pneumoniae* Hu15 plasmid pSpn / *Streptococcus pneumoniae* Hu17 plasmid pSpn / *Streptococcus pneumoniae* pSpnP1 plasmid
	>spacer40	
	GGGTATCATTGTATCTAGTGATGGACCTGA	
	>spacer41	
	ATTTGAAAAATGCACAACAGCGTTTGATAG	*Streptococcus* phage TP-J34
CS8 (287797..290540) (rev_com)	>spacer1	
	AGAACGTATTCCAAAACCTCTTTACGATTA	*Streptococcus* phages: V2 / R1 / MM25 / 31B4 / 16B8 / 9B4 / 9A / 7T / P9854 / P9853 / P9852 / P9851 / P8922 / P8921 / P7954 / P7953 / P7952 / P7633 / P7632 / P7631 / P7602 / P7601 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P3684 / P3681 / 128 / 53 / CHPC577 / ALQ13.2 / vB_SthS_VA460 / vB_SthS_VA698; *Streptococcus thermophilus* bacteriophages: kappa3 / 2972 / DT1
	>spacer2	
	TTAACTGTTATCAAAATGATAAGATAGTCT	
	>spacer3	
	CGTTGATGTTTATTCAAGTAAAATAATTAA	*Streptococcus thermophilus* temperate bacteriophage O1205
	>spacer4	
	TCCTTTCACGGGTAGCACACTAACATACAC	
	>spacer5	
	GTTGGCAATGCAAACAACCTTTATGAACCG	*Streptococcus* phage TP-J34; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer6	
	TTTATTTCCTTGCGATAACGTTCCACCTTT	*Streptococcus* phage TP-J34; *Streptococcus thermophilus* bacteriophages: Sfi11 / O1205
	>spacer7	
	AGATTATAAGGAACACAACCAACTATATAG	

	>spacer8	
	ACGACATCAAGCTGATTGTCTTCTACATAA	*Streptococcus* phages: STP1 / B5 / P9902 / 73; *Streptococcus thermophilus* bacteriophage DT1
	>spacer9	
	TTTGGAATACTGAATGTTTTACTGAAAATC	
	>spacer10	
	ACACCACTATCTTTTCCTCCTGAAAATGAA	
	>spacer11	
	GTAATTCCACGAAATTATCAACCTTATGCA	*Streptococcus* phage P9903 / P5641 / P4761 / D4276 / 73; *Streptococcus thermophilus* bacteriophages: DT1 / Sfi21
	>spacer12	
	TTGGAGGATTGCCCCATATTCCCAAGAGT	
	>spacer13	
	GAGAGGCGTTAAATATAGAAATGCAAGATT	*Streptococcus* phage vB_SthS_VA698; *Streptococcus thermophilus* temperate bacteriophage O1205
	>spacer14	
	TTTTAACGTCATCAGTCCACCGCCTTAAAT	
	>spacer15	
	CACCTCTTTCGATGGAAAGGTATCCTTCTA	*Streptococcus* phage vB_SthS_VA460
	>spacer16	
	GACCAAAGTTTGATTATAGAGCTATACACC	
	>spacer17	
	ACCATCATTCTTACCATTACAACTGTAATG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer18	
	ATACGAATTCGGTTCGCACAATTACAATTC	
	>spacer19	
	TATCAACGCAATCATTACAACAACTTCAAACA	
	>spacer20	
	ATCTACGTGTCAATACATATCACAAAACAG	*Streptococcus* phages: P7954 / pC2S4; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer21	
	ATTTTTAGAAATTTCTGATATAATAATGA	
	>spacer22	

	TTGTTGGAACAAGGACGACTTGGTAAACTA	*Streptococcus* virus 9874
	>spacer23	
	CATATTAAGCTGACTGGGCCTAATGCTTTT	*Streptococcus* phages: P9854 / P9852 / P9851 / 858 / TP-778L; *Lactococcus* phages: AM5 / AM4; *Streptococcus thermophilus* bacteriophages: 2972 / O1205
	>spacer24	
	TTCATAGCATACCGTAGTTGTAAAATCTAT	
	>spacer25	
	AACATTTAGGGAATGAAATTGATAAGACTG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer26	
	AACATGAGAAACTGTAGAAAACAAGCAATA	
	>spacer27	
	TGGTGAAGATGGCAGTCATAAATGGCACATT	*Streptococcus* phages: vB_SthS_VA214 / P9851 / P0091
	>spacer28	
	AAGGGTTGAAAAATGTTGGTATATCAAACG	*Streptococcus* phages: V2 / R1 / L5A1 / C0 / 31B4 / 9B4 / A0 / P7633 / P7632 / P7631 / P7574
	>spacer29	
	TTCTGGTAGTGGATTTAGTCAAACAGATGT	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer30	
	TCCATAGAGCGTCTTAAACAAAGAATAGTC	
	>spacer31	
	GTTGGCAATGCAAACAACCTTTATGAACCG	*Streptococcus* phage TP-J34; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer32	
	CATATTAAGCTGACTGGGCCTAATGCTTTT	*Streptococcus* phages: P9854 / P9852 / P9851 / 858 / TP-778L; *Lactococcus* phages: AM5 / AM4; *Streptococcus thermophilus* bacteriophages: 2972 / O1205
	>spacer33	
	TTCATAGCATACCGTAGTTGTAAAATCTAT	
	>spacer34	
	AACATTTAGGGAATGAAATTGATAAGACTG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19

	>spacer35	
	AACATGAGAAACTGTAGAAAACAAGCAATA	
	>spacer36	
	TGGTGAAGATGGCAGTCATAAATGGCACATT	*Streptococcus* phages: vB_SthS_VA214 / P9851 / P0091
	>spacer37	
	TTATGATTGAATGACATGGTTGTATAAGTA	*Streptococcus* phages: P9854 / P9852 / P9851 / TP-778L / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer38	
	TTTCTTTAGGAATACCAGGGAGTTCAGCTT	*Enterococcus faecium* HB6 plasmid pAR6 / *Enterococcus faecium* JH95 plasmid pJS33 / *Enterococcus faecium* plasmid pEFR
	>spacer39	
	TGGCAGAGATTACACAGCAACGGAAACAGC	*Streptococcus pneumoniae* Hu15 plasmid pSpn / *Streptococcus pneumoniae* Hu17 plasmid pSpn / *Streptococcus pneumoniae* pSpnP1 plasmid
	>spacer40	
	GGGTATCATTGTATCTAGTGATGGACCTGA	
	>spacer41	
	ATTTGAAAAATGCACAACAGCGTTTGATAG	*Streptococcus* phage TP-J34
LMD-9 (649125..650217)	>spacer1	
	ATGATGATGAAGTATCGTCATCTACTAAC	
	>spacer2	
	CTTCACCTCAAATCTTAGAGCTGGACTAAA	*Streptococcus* phages: P7602 / P3684; *Streptococcus thermophilus* bacteriophage 7201
	>spacer3	
	ATGTCTGAAAAATAACCGACCATCATTACT	*Streptococcus* phages: 9A / P8921 / P3684 / P3681 / TP-778L
	>spacer4	
	GAAGCTCATCATGTTAAGGCTAAAACCTAT	*Streptococcus* phage vB_SthS_VA460 / vB_SthS_VA698 / P7133 / P3684 / P3681 / 128
	>spacer5	
	TAGTCTAAATAGATTTCTTGCACCATTGTA	*Streptococcus* phage vB_SthS_VA460
	>spacer6	

	ATTCGTGAAAAAATATCGTGAAATAGGCAA	*Streptococcus suis* strain 6936 plasmid unnamed1 / *Streptococcus suis* strain ST104 zonular occludens toxin (zot) gene
	>spacer7	
	TCTAGGCTCATCTAAAGATAAATCAGTAGC	*Streptococcus* phages: vB_SthS_VA460 / vB_SthS_VA214 / P9853 / P7955 / P7573 / P7572 / P7571 / P4761 / D4276 / CHPC926 / CHPC577 / 9874 / 9873 / 9872 / ALQ13.2 / 858; *Streptococcus thermophilus* bacteriophages: J1 / ST3 / Sfi16A / S3b
	>spacer8	
	TAAAAACATGGGGCGGCGGTAATAGTGTAAG	
	>spacer9	
	ACAACCAGCAAAGAGAGCGCCGACAACATT	
	>spacer10	
	TATAACACAGGTTTAGAGGATGTTATACTT	
	>spacer11	
	CTAGAAGCTCAAGCGGTAAAAGTTGATGGCG	
	>spacer12	
	CTTTGAGGGCAAGCCCTCGCCGTTCCATTT	
	>spacer13	
	AACTACCAAGCAAATCAGCAATCAATAAGT	*Streptococcus* phages: 16B8 / 7T / P7633 / P7632 / P7602 / P7573 / P0091; *Streptococcus thermophilus* bacteriophage Sfi19
	>spacer14	
	CTATAAGTGACAATCAGCGTAGGGAATACG	
	>spacer15	
	ATCAGTGCGGTATATTTACCCTAGACGCTA	
	>spacer16	
	AACAGTTACTATTAATCACGATTCCAACGG	*Streptococcus* phage vB_SthS_VA460
SMQ-301 (649756..650847)	>spacer1	
	CGTGGCAGCGTGGTCGGGTTTAATAGCCCG	*Streptococcus* phages: vB_SthS_VA214 / P9903 / P9902 / P8921 / P7632 / P7631 / P7574 / P7152 / P5651 / P0091 / 128
	>spacer2	
	AAGCCCAAGTCAGAGCATCCGTCCAAGCC	*Streptococcus suis* strain 6936 plasmid unnamed1
	>spacer3	

	ATTGGGTTTCGGTAAGAACTAAACATACCA	*Streptococcus* phage YMC-2011
	>spacer4	
	CACAAAATAATTCGGTAGTTTTTACTAACT	*Streptococcus* phages: P9854 / P9851
	>spacer5	
	TTTGACCGTTTATTTAGACGTGCTAAAGT	
	>spacer6	
	AATATCTACAGGTCACTACAAAGCTACGCT	*Streptococcus* phages: vB_SthS_VA460 / vB_SthS_VA698 / P9854 / P9851 / P7573 / P5641 / P3684 / P3681; *Streptococcus thermophilus* bacteriophages: 7201 / Sfi21 / Sfi19
	>spacer7	
	TCTAGGCTCATCTAAAGATAAATCAGTAGC	*Streptococcus* phages: vB_SthS_VA460 / vB_SthS_VA214 / P9853 / P7955 / P7573 / P7572 / P7571 / P4761 / D4276 / CHPC926 / CHPC577 / 9874 / 9873 / 9872 / ALQ13.2 / 858; *Streptococcus thermophilus* bacteriophages: J1 / ST3 / Sfi16A / S3b
	>spacer8	
	TAAAAACATGGGGCGGCGGTAATAGTGTAAG	
	>spacer9	
	ACAACCAGCAAAGAGAGCGCCGACAACATT	
	>spacer10	
	TATAACACAGGTTTAGAGGATGTTATACTT	
	>spacer11	
	CTAGAAGCTCAAGCGGTAAAAGTTGATGGCG	
	>spacer12	
	CTTTGAGGGCAAGCCCTCGCCGTTCCATTT	
	>spacer13	
	AACTACCAAGCAAATCAGCAATCAATAAGT	*Streptococcus* phages: 16B8 / 7T / P7633 / P7632 / P7602 / P7573 / P0091; *Streptococcus thermophilus* bacteriophage Sfi19
	>spacer14	
	CTATAAGTGACAATCAGCGTAGGGAATACG	
	>spacer15	
	ATCAGTGCGGTATATTTACCCTAGACGCTA	
	>spacer16	

	AACAGTTACTATTAATCACGATTCCAACGG	*Streptococcus* phage vB_SthS_VA460
GABA (652778..655449)	>spacer1	
	GACTCTATTGGCGTTTTACCACTTATTGCT	
	>spacer2	
	ATTCTAGATAGTAACGGAAACAAAGTACAG	
	>spacer3	
	GTATCCAGTTATCGTCATTCTCTTTACCTG	*Streptococcus* phages: P4761 / TP-778L
	>spacer4	
	CCAAATTTGCATTAAACAAAACGCTCCTTC	*Streptococcus* phages: vB_SthS_VA214 / P9901 / P9854 / P9851 / P7574 / P7573 / P0091
	>spacer5	
	CTCGTAACTAAGTTTGAAAACTTCTTTGAC	*Streptococcus* phages: STP2 / STP1 / P7955 / P7951 / P7602 / P7573 / P7132 / P5641 / P0092 / D4276 / CHPC926 / 9874
	>spacer6	
	ACAGTAAGATACACGTAGTTGATGAATTG	*Streptococcus* phages: P7955 / P7951 / P7572 / P7571 / TP-778L; *Streptococcus thermophilus* bacteriophage O1205
	>spacer7	
	AAGACATATCTTTTTAACATCATGAAGACG	*Streptococcus* phages: P9853 / P9852 / P7951 / P7572 / P7571 / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer8	
	TAAAACTTACTGCAGACAGTGTTAATTATC	
	>spacer9	
	CACAAAATAATTCGGTAGTTTTTACTAACT	*Streptococcus* phages: P9854 / P9851
	>spacer10	
	TTTGACCGTTTATTTAGACGTGCTAAAGT	
	>spacer11	
	AATATCTACAGGTCACTACAAAGCTACGCT	*Streptococcus* phages: vB_SthS_VA460 / vB_SthS_VA698 / P9854 / P9851 / P7573 / P5641 / P3684 / P3681; *Streptococcus thermophilus* bacteriophages: 7201 / Sfi21 / Sfi19
	>spacer12	
	TGATAGAGAATGGTCGTTAAATGGTGTAAT	*Streptococcus* phages: P7632 / P7631 / P7573 / P7134 / P3684 / P3681 / P7602

	>spacer13	
	TCAGTTTCAAATGCTCCACCACCGATATTA	*Streptococcus* phages: P3684 / P3681
	>spacer14	
	GTTCGGGTTCAGGTGGTTTTATACTCAACT	*Streptococcus* phage D4276
	>spacer15	
	AACATGCATCAGTTTGGATACGAGGTCAAG	*Streptococcus* phage P0091
	>spacer16	
	GAACAAGTAGAAGGAGGAGTGGAAGAGTCG	
	>spacer17	
	ACGAGTGGGCTATCAGTCACGGACTGGATA	*Streptococcus* phages: MM25 / P9852 / P9851 / P8922 / P8921 / P7955 / P7951 / P7633 / P7632 / P7631 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / D4276 / 128 / Abc2 / ALQ13.2; *Streptococcus thermophilus* bacteriophages: Sfi11 / DT1 / SFi18 / Sfi21 / Sfi19
	>spacer18	
	CTTGACTAATACGTCTGGATTTACTGCACT	*Streptococcus* phages: V2 / R1 / M19 / C0 / 16B8 / 9B4 / 7T / P9901 / P9854 / P7574 / P7573 / P3684 / P3681; *Streptococcus thermophilus* bacteriophage 7201
	>spacer19	
	AAATAGACTGGAATGGATTGAGAATAATCT	*Streptococcus* phages: P0095 / 5093
	>spacer20	
	TTGAAAAAGATAGACAATATCAAACTCTGG	
	>spacer21	
	TCAGAAAGCGGTTGCTGTGCCTGTGCTGG	
	>spacer22	
	AGACCATCCGTTCTTTGTGTTTAGAGCCAT	
	>spacer23	
	GATACCGTTTTTACAATAATATCTATACCA	*Streptococcus* phage vB_SthS_VA214
	>spacer24	
	GGCCGCATCGGAAAATGGGCTTGTTATCGG	
	>spacer25	
	CTGCTCGATAAATACTTAGCTACTGGTTTT	
	>spacer26	

	TGTTGATAGCGTCACCCACTGAGTTATCTTG	
	>spacer27	
	GGAAACGCTAGCTTAGCTAAATTTCAAGCA	*Streptococcus* phages: vB_SthS_VA214 / 5093
	>spacer28	
	CTTGAATCTGCTTGACTGATAATGTGCTAT	*Streptococcus* phages: vB_SthS_VA460 / P8921 / P3684 / P3681; *Streptococcus thermophilus* bacteriophage 7201
	>spacer29	
	GGAAACGCTAGCTTAGCTAAATTTCAAGCA	*Streptococcus* phages: vB_SthS_VA214 / 5093
	>spacer30	
	CTTGAATCTGCTTGACTGATAATGTGCTAT	*Streptococcus* phages: vB_SthS_VA460 / P8921 / P3684 / P3681; *Streptococcus thermophilus* bacteriophage 7201
	>spacer31	
	TCATAGCTACTTTGAGAGTGAAGAAGCACA	*Streptococcus* phages: P7955 / P7951 / CHPC926 / 9873 / 9872 / 9871
	>spacer32	
	AAACCGACTTGAGTGGATTGAAAATAATCT	*Streptococcus* phage 9874; *Streptococcus thermophilus* bacteriophage DT1
	>spacer33	
	ATTGATAGAACCGGTGAAATTTTTCCTATC	*Streptococcus* phage vB_SthS_VA214
	>spacer34	
	TAGATGACATAAAAATTTCACACAGTGAAG	*Streptococcus* phages: P7955 / P7573; *Streptococcus thermophilus* bacteriophage 7201
	>spacer35	
	CGAACATTGCAGAACTGGGGAAATGGGTAC	
	>spacer36	
	TAAGAAACCTGTCGGAGAAGCTAAAAGCTG	*Streptococcus* phage MM25 / P9901 / P8921 / P7601 / P7152 / P7132 / D4276 / 53
	>spacer37	
	CTCACTAGTCCATTAGTGCGGTTAGGGAA	*Streptococcus* phage 9A / P8922 / P8921 / P7953 / P7633 / P7632 / P7631 / P7572 / P7571 / P7154 / P7151 / P7132 / P5652 / P5651 / 53; *Streptococcus thermoph*ilus bacteriophage DT1
	>spacer38	

	TGGTACTACACCTCAAACGGTGAGTGCATG	*Streptococcus* phage MM25 / 9A / vB_SthS_VA698 / P8922 / P8921 / P7954 / P7953 / P7952 / P7633 / P7632 / P7631 / P7154 / P7152 / P7151 / P7134 / P7133 / P5652 / P5651 / D4276; *Streptococcus thermophilus* bacteriophages: DT1 / 7201
	>spacer39	
	GCAACAATGGTTATAATTATCCCACACAGG	
	>spacer40	
	GCATACGCTGTTACAAACACACCCCACCC	
ST3 (652775..653929)	>spacer1	
	GACGTTCAATGAAATATTCTTATCAGACAT	*Streptococcus pyogenes* A996 plasmid pA996
	>spacer2	
	TCTGACGATAAAGAGAATATCTCAAAGGGT	
	>spacer3	
	GGGTGCGATGGATTCAATGACAAAGAGCAC	*Streptococcus* phages: M19 / 9A / P8922 / P8921 / P7633 / P7632 / P7631 / P7154 / P7152 / P7151 / P5651 / 128 /
	>spacer4	
	TAAAACTTACTGCAGACAGTGTTAATTATC	
	>spacer5	
	CACAAAATAATTCGGTAGTTTTTACTAACT	*Streptococcus* phages: P9854 / P9851
	>spacer6	
	TTTGACCGTTTATTTAGACGTGCTAAAGT	
	>spacer7	
	AATATCTACAGGTCACTACAAAGCTACGCT	*Streptococcus* phages: vB_SthS_VA460 / vB_SthS_VA698 / P9854 / P9851 / P7573 / P5641 / P3684 / P3681; *Streptococcus thermophilus* bacteriophages: 7201 / Sfi21 / Sfi19
	>spacer8	
	TGATAGAGAATGGTCGTTAAATGGTGTAAT	*Streptococcus* phages: P7632 / P7631 / P7573 / P7134 / P3684 / P3681 / P7602
	>spacer9	
	TCAGTTTCAAATGCTCCACCACCGATATTA	*Streptococcus* phages: P3684 / P3681
	>spacer10	
	GTTCGGGTTCAGGTGGTTTTATACTCAACT	*Streptococcus* phage D4276
	>spacer11	

	AACATGCATCAGTTTGGATACGAGGTCAAG	*Streptococcus* phage P0091
	>spacer12	
	CGAACATTGCAGAACTGGGGAAATGGGTAC	
	>spacer13	
	TAAGAAACCTGTCGGAGAAGCTAAAAGCTG	*Streptococcus* phage MM25 / P9901 / P8921 / P7601 / P7152 / P7132 / D4276 / 53
	>spacer14	
	CTCACTAGTCCATTAGTGCGGTTAGGGAA	*Streptococcus* phage 9A / P8922 / P8921 / P7953 / P7633 / P7632 / P7631 / P7572 / P7571 / P7154 / P7151 / P7132 / P5652 / P5651 / 53; *Streptococcus thermoph*ilus bacteriophage DT1
	>spacer15	
	TGGTACTACACCTCAAACGGTGAGTGCATG	*Streptococcus* phage MM25 / 9A / vB_SthS_VA698 / P8922 / P8921 / P7954 / P7953 / P7952 / P7633 / P7632 / P7631 / P7154 / P7152 / P7151 / P7134 / P7133 / P5652 / P5651 / D4276; *Streptococcus thermophilus* bacteriophages: DT1 / 7201
	>spacer16	
	GCAACAATGGTTATAATTATCCCACACAGG	
	>spacer17	
	GCATACGCTGTTACAAACACACCCCACCC	
ND03 (639512..641921)	>spacer1	
	TCTTTCTATAGCATAATTGTTAGTTGTCCT	*Streptococcus* phages: P7951 / D4276
	>spacer2	
	GAAAGAATCGGTCTTCTAGATGGATTCCAA	*Streptococcus* phages: vB_SthS_VA214 / STP2 / STP1 / L5A1 / C0 / B5 / B0 / 7A5 / A0 / P0094 / P0093 / P0091 / CHPC1151 / CHPC926 / 9871 / Abc2
	>spacer3	
	TAGGAAATTCATACCCTGCAGATATATCAT	
	>spacer4	
	AACGATAAAAAAGAAGAAGCTTTCACGCCT	
	>spacer5	
	TCGATTTTAAGCCGTGTATTTTGCTAGGTA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer6	
	GCTTGGAGGCGTCAATGTTCGCAGCTTATC	

	>spacer7	
	ACCTGTCATCTCTGGGAGTTAAATTTCCTT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer8	
	ATTAGCACGAGCCTCTGTATACTGTTCTTT	*Streptococcus* phages: R1 / MM25 / L5A1 / 31B4 / 7A5 / P9903 / P9902 / P8921 / P7602 / P7601 / P7574 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P5641 / D4276 / 53 / 73 / Abc2 / vB_SthS_VA460; *Streptococcus thermophil*us bacteriophage DT1
	>spacer9	
	TTTGGTCTGGTGCGCCTTTTATATTGAATT	
	>spacer10	
	GGTTCCCTATGAGCAAAGGACAACGCTTT	
	>spacer11	
	TCGATTTTAAGCCGTGTATTTTGCTAGGTA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer12	
	GCTTGGAGGCGTCAATGTTCGCAGCTTATC	
	>spacer13	
	ACATTGAGTAAGTCAATAGTTTGCGATGGC	*Streptococcus* phage 9871
	>spacer14	
	TAGAATAAGTGAATGGGAAACTGACGGAAC	
	>spacer15	
	ACATAGAAGGCTAAGCAATCAAGCAAAAGA	
	>spacer16	
	ACGACATCAAGCTGATTATCTTCTACATAC	*Streptococcus* phages: V2 / R1 / MM25 / 31B4 / 16B8 / 9B4 / 7A5 / 7T / P7133 / P5652 / P0091 / Abc2 / vB_SthS_VA460 / vB_SthS_VA214
	>spacer17	
	CTGTCTAACATAGTCCCAAACTTTATCATA	
	>spacer18	
	CTAAAAACGGTGTTCTATATCGAGGTCAAC	
	>spacer19	
	TGTTTCCATCCGTGGTGAAATGGTAGAAGT	*Streptococcus* phages: TP-778L / 858
	>spacer20	

	TTGAGCATTTACTTACTAATTTCAAACGGT	
	>spacer21	
	AGATGACCTATAATATAATCAAAATCACCT	
	>spacer22	
	TCTATCTTAGGCGGTGGACTGATGACGTTA	
	>spacer23	
	GATAGATATATATATTATATCACAATCCTA	*Streptococcus* phage P0091
	>spacer24	
	TTATAGTATATCTAATGTGTTCCCGTTTGT	*Streptococcus* phages: L5A1 / C0 / B5 / B0 / 7A5 / A0 / P7574 / P7572 / P7571 / P7132 / 53 / ALQ13.2
	>spacer25	
	TTGAAATTTATAATATGGATGTCGATGATG	*Streptococcus* phage P7573; *Streptococcus thermophilus* bacteriophage O1205
	>spacer26	
	TCTAGAGACCTCTTCGTGTTGTCACAACCT	*Streptococcus* phage vB_SthS_VA214 / vB_SthS_VA698 / V2 / R1 / M19 / L5A1 / C0 / 16B8 / 9B4 / 9A / A0 / 7T / P9903 / P9902 / P9901 / P8922 / P8921 / P7632 / P7631 / P7154 / P7152 / P7151 / P7134 / P7133 / P5652 / P5651 / P5641 / P3684 / P0091 / 128; *Streptococcus thermophilus* bacteriophage 7201
	>spacer27	
	GCTCTTTCCTACTCCTGTGTCTCCTATTA	*Streptococcus* phage P0092
	>spacer28	
	GTTTATCACTTATTTGATATACACTTATAA	
	>spacer29	
	ATATCGGTGGTGGTGCGTTTGAAACTGACA	*Streptococcus* phage vB_SthS_VA698 / V2 / R1 / M19 / C0 / 16B8 / 9B4 / 7T / P9901 / P7574; *Streptococcus thermophilus* bacteriophage 7201
	>spacer30	
	GACCCACGGATACCGCCAAAACAGCACGTG	*Streptococcus* phages: P7574 / P7573
	>spacer31	
	GAAGTGATAGACGTTGTTAACTCATCAGCA	
	>spacer32	

	TTTCTGTGTACTCCATCGGCTTTTGAGTGT	*Streptococcus* phage CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871; *Lactococcus* phage P335
	>spacer33	
	GCTTTGCGAGTAACGCCGGTTGCTTTTCCT	
	>spacer34	
	TTGGCATCAGTGAATTTGATGCCTGGAAGT	*Streptococcus* phage P5641
	>spacer35	
	AATTTTGCCCCTTCTTTGCCCCTTGACTAG	*Streptococcus* phage 5093
	>spacer36	
	ACCATTAGCAATCATTTGTGCCCATTGAGT	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
APC151 (1689194.. 1691603)	>spacer1	
	TCTTTCTATAGCATAATTGTTAGTTGTCCT	*Streptococcus* phages: P7951 / D4276
	>spacer2	
	GAAAGAATCGGTCTTCTAGATGGATTCCAA	*Streptococcus* phages: vB_SthS_VA214 / STP2 / STP1 / L5A1 / C0 / B5 / B0 / 7A5 / A0 / P0094 / P0093 / P0091 / CHPC1151 / CHPC926 / 9871 / Abc2
	>spacer3	
	TAGGAAATTCATACCCTGCAGATATATCAT	
	>spacer4	
	AACGATAAAAAAGAAGAAGCTTTCACGCCT	
	>spacer5	
	TCGATTTTAAGCCGTGTATTTTGCTAGGTA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer6	
	GCTTGGAGGCGTCAATGTTCGCAGCTTATC	
	>spacer7	
	ACCTGTCATCTCTGGGAGTTAAATTTCCTT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer8	
	ATTAGCACGAGCCTCTGTATACTGTTCTTT	*Streptococcus* phages: R1 / MM25 / L5A1 / 31B4 / 7A5 / P9903 / P9902 / P8921 / P7602 / P7601 / P7574 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P5641 / D4276 / 53 / 73 / Abc2 / vB_SthS_VA460; *Streptococcus thermophil*us bacteriophage DT1

	>spacer9	
	TTTGGTCTGGTGCGCCTTTTATATTGAATT	
	>spacer10	
	GGTTCCCTATGAGCAAAGGACAACGCTTT	
	>spacer11	
	TCGATTTTAAGCCGTGTATTTTGCTAGGTA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer12	
	GCTTGGAGGCGTCAATGTTCGCAGCTTATC	
	>spacer13	
	ACATTGAGTAAGTCAATAGTTTGCGATGGC	*Streptococcus* phage 9871
	>spacer14	
	TAGAATAAGTGAATGGGAAACTGACGGAAC	
	>spacer15	
	ACATAGAAGGCTAAGCAATCAAGCAAAAGA	
	>spacer16	
	ACGACATCAAGCTGATTATCTTCTACATAC	*Streptococcus* phages: V2 / R1 / MM25 / 31B4 / 16B8 / 9B4 / 7A5 / 7T / P7133 / P5652 / P0091 / Abc2 / vB_SthS_VA460 / vB_SthS_VA214
	>spacer17	
	CTGTCTAACATAGTCCCAAACTTTATCATA	
	>spacer18	
	CTAAAAACGGTGTTCTATATCGAGGTCAAC	
	>spacer19	
	TGTTTCCATCCGTGGTGAAATGGTAGAAGT	*Streptococcus* phages: TP-778L / 858
	>spacer20	
	TTGAGCATTTACTTACTAATTTCAAACGGT	
	>spacer21	
	AGATGACCTATAATATAATCAAAATCACCT	
	>spacer22	
	TCTATCTTAGGCGGTGGACTGATGACGTTA	
	>spacer23	
	GATAGATATATATATTATATCACAATCCTA	*Streptococcus* phage P0091
	>spacer24	

	TTATAGTATATCTAATGTGTTCCCGTTTGT	*Streptococcus* phages: L5A1 / C0 / B5 / B0 / 7A5 / A0 / P7574 / P7572 / P7571 / P7132 / 53 / ALQ13.2
	>spacer25	
	TTGAAATTTATAATATGGATGTCGATGATG	*Streptococcus* phage P7573; *Streptococcus thermophilus* bacteriophage O1205
	>spacer26	
	TCTAGAGACCTCTTCGTGTTGTCACAACCT	*Streptococcus* phage vB_SthS_VA214 / vB_SthS_VA698 / V2 / R1 / M19 / L5A1 / C0 / 16B8 / 9B4 / 9A / A0 / 7T / P9903 / P9902 / P9901 / P8922 / P8921 / P7632 / P7631 / P7154 / P7152 / P7151 / P7134 / P7133 / P5652 / P5651 / P5641 / P3684 / P0091 / 128; *Streptococcus thermophilus* bacteriophage 7201
	>spacer27	
	GCTCTTTCCTACTCCTGTGTCTCCTATTA	*Streptococcus* phage P0092
	>spacer28	
	GTTTATCACTTATTTGATATACACTTATAA	
	>spacer29	
	ATATCGGTGGTGGTGCGTTTGAAACTGACA	*Streptococcus* phage vB_SthS_VA698 / V2 / R1 / M19 / C0 / 16B8 / 9B4 / 7T / P9901 / P7574; *Streptococcus thermophilus* bacteriophage 7201
	>spacer30	
	GACCCACGGATACCGCCAAAACAGCACGTG	*Streptococcus* phages: P7574 / P7573
	>spacer31	
	GAAGTGATAGACGTTGTTAACTCATCAGCA	
	>spacer32	
	TTTCTGTGTACTCCATCGGCTTTTGAGTGT	*Streptococcus* phage CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871; *Lactococcus* phage P335
	>spacer33	
	GCTTTGCGAGTAACGCCGGTTGCTTTTCCT	
	>spacer34	
	TTGGCATCAGTGAATTTGATGCCTGGAAGT	*Streptococcus* phage P5641
	>spacer35	
	AATTTTGCCCCTTCTTTGCCCCTTGACTAG	*Streptococcus* phage 5093
	>spacer36	

	ACCATTAGCAATCATTTGTGCCCATTGAGT	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
JIM 8232 (712332..715137)	>spacer1	
	ATGAGGTCGAGAAGGAAAAACGATATGCAG	*Streptococcus* phage M19
	>spacer2	
	ATCCTCAATAGCATTGACAAAGTCTTTTCT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer3	
	TCAATCTTCCTGTCTTTGAAATATTTCTTT	*Streptococcus* phage 5093
	>spacer4	
	TTGGAATTTAACGTTTGAAGAACGGAAGAA	*Streptococcus* phage vB_SthS_VA460
	>spacer5	
	GTGGTCTTTCGTGGGTGCTGCGCTCTGGTT	
	>spacer6	
	ACTAAAATCATTTTCAACGAGTTGAGAGAC	*Streptococcus* phages: P7602 / P7574; *Streptococcus thermophilus* bacteriophage 7201 (3rd hit)
	>spacer7	
	TATATTCGCGAGGGTGATATAGTCGTGGTG	*Lactobacillus salivarius* CECT 5713 plasmid pHN1
	>spacer8	
	CATATATGATAGTTTGTCAACACCTTTTAT	*Streptococcus* phages: P7602 / P5641 / CHPC577
	>spacer9	
	TAAATGGGTTATGCTGTTCAATATGCGTCC	*Streptococcus* phage P7573
	>spacer10	
	AAACTTGCCAGCAATCGGAATGTAAAGGCC	*Streptococcus thermophilus* bacteriophage DT2
	>spacer11	
	TTAGACCCTAAAGAGGTAGAGCGCCATATA	
	>spacer12	
	AAAAGGTTGGCTAGTGTAATAAGTTCGTCA	*Streptococcus* phages: M19 / L5A1 / C0 / B5 / B0 / 7A5 / A0 / P7955 / P7573 / P7132 / 53 / ALQ13.2
	>spacer13	
	TACCATTGTTTCTCTCCACCGTGGAAGGTT	*Streptococcus* phages: P0094 / P0093 / P0091 / CHPC1151
	>spacer14	
	GGCATTAGACAGATACATTGATAGCCAACT	

	>spacer15	
	AAGAAGCAGCAAGATGCATCAAACACCGCA	*Streptococcus* phage V2 / STP2 / R1 / M19 / L5A1 / C0 / B5 / B0 / 31B4 / 9B4 / A0 / 7T / vB_SthS_VA698 / P9902 / P9901 / P7951 / P7573 / P7151 / P7134 / P3684 / P3681 / 53 / Abc2 / ALQ13.2
	>spacer16	
	TGAAAATCTCACTAGTCCATTAGTGCGGTT	
	>spacer17	
	ATCACCCAACTTTCTTGTGCTTTCTAAGTG	*Streptococcus* phages: V2 / STP2 / STP1 / R1 / L5A1 / C0 / B5 / B0 / 31B4 / 16B8 / 9B4 / 9A / 7A5 / A0 / 7T / vB_SthS_VA698 / P9903 / P9902 / P8922 / P8921 / P7633 / P7632 / P7631 / P7601 / P7574 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / D4276 / 128 / 53 / 73 / Abc2; *Streptococcus thermophilus* bacteriophage DT1
	>spacer18	
	GCTTAGCTGTCCAATCCACAAACGTGGATG	*Streptococcus* phages: V2 / R1 / M19 / C0 / 31B4 / 16B8 / 9B4 / 9A / 7A5 / A0 / 7T / vB_SthS_VA698 / P9902 / P8922 / P8921 / P7633 / P7632 / P7631 / P7601 / P7154 / P7152 / P7151 / P7132 / P5652 / P5651 / P5641 / P3684 / P3681 / 128 / 53 / 73
	>spacer19	
	GCTAACTTAAGCAAGCAGTACGAAAAGCAG	*Streptococcus thermophilus* bacteriophage 7201
	>spacer20	
	AAAAAAGATGGTATCAATTATCTAGAATTG	
	>spacer21	
	CTCTTAAGGTCATTTCAGTGTCTTGGTTA	*Streptococcus* phage M19; *Streptococcus thermophilus* bacteriophage 7201
	>spacer22	
	ATTGTCTATTACGACAACATGGAAGATAGT	*Streptococcus* phages: STP2 / STP1 / L5A1 / C0 / B5 / B0 / 7A5 / A0 / P9853 / P5641 / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer23	
	TACACTAGTAGTTGTGTCCCAAATTTTAGA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer24	

	ACTAAAATCATTTTCAACGAGTTGAGAGAC	*Streptococcus* phages: P7602 / P7574; *Streptococcus thermophilus* bacteriophage 7201 (3rd hit)
	>spacer25	
	TATATTCGCGAGGGTGATATAGTCGTGGTG	*Lactobacillus salivarius* CECT 5713 plasmid pHN1
	>spacer26	
	TAGGGATATGAAGCCCAAATTGTTGCAATG	*Streptococcus* phage D4276
	>spacer27	
	TTTCCCGGTGTATAGATTGGCTCTATAGTCT	*Streptococcus* phages: P9903 / P9902 / P9901 / 73 / Abc2
	>spacer28	
	TATATATAAGTTTCCGAAAGTAGACGAACC	*Streptococcus* phages: M19 / Abc2
	>spacer29	
	CATTATTAAACCCATGCTTTTTACCACACT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer30	
	CATTATTAAACCCATGCTTTTTACCACACT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer31	
	AGCCGATAGTTAAATCAACTGATTGTGGAT	
	>spacer32	
	CGTGCTACAAATCTCTTTACGACAATTTAG	*Streptococcus thermophilus* bacteriophage 7201
	>spacer33	
	TTTTTTTGATACATTTAAGGGCGCTTTCAT	*Streptococcus* phage CHPC1151
	>spacer34	
	ACGAGAGTCTTTCAGTCTATCCATTGCCTG	*Streptococcus* phages: P0095 / P0094 / P0093 / P0092 / CHPC1151 / 5093
	>spacer35	
	CATTTTCCGTTTTTTTTGTCTCTACCCCAT	
	>spacer36	
	AGCTGTGATAACACGAAGAGAACTCTAGA	*Streptococcus* phages: V2 / R1 / M19 / L5A1 / C0 / 16B8 / 9B4 / 9A / A0 / 7T / vB_SthS_VA214 / vB_SthS_VA698 / P9903 / P9902 / P9901 / P8922 / P8921 / P7632 / P7631 / P7154 / P7152 / P7151 / P7134 / P7133 / P5652 / P5651 / P5641 / P3684 /

		P0091 / 128; S*treptococcus thermophilus* bacteriophage 7201
	>spacer37	
	AAACATTGATGAGTATTATTGAAAAACTTCA	*Streptococcus* phages: V2 / STP2 / R1 / M19 / L5A1 / C0 / B5 / B0 / 31B4 / 9B4 / A0 / 7T / vB_SthS_VA698 / P9902 / P7573 / P7151 / P7134 / P3684 / P3681 / 53 / Abc2 / ALQ13.2
	>spacer38	
	GGGTCAAGGTATAAGATAGAGTTTTCTATA	
	>spacer39	
	TAGGTACTTTAACCTGACCAGATACAATC	
	>spacer40	
	CCTCAAAGCTTAAAATTGGGCTGAAGTAGA	*Streptococcus* phages: M19 / P5641
	>spacer41	
	AGGTTGTAACAAAAAAGATGCTAACAAGA	
	>spacer42	
	TAATGCTACATCTCAAAGGATGATCCCAGA	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
MN-ZLW-002 (630165..632177)	>spacer1	
	GATAGCAATAGCTTTCTTGACCTAAAAGAC	*Streptococcus thermophilus* bacteriophage Sfi21
	>spacer2	
	GAGGTCTGTAATTTCATTCCCTCGTAATCT	*Streptococcus suis* 6936 plasmid unnamed1
	>spacer3	
	AAAGGTTTCTCTAAACACATGCGGAATAT	
	>spacer4	
	GTCATAGTACCAAGCACAAATAACGTTAGT	
	>spacer5	
	GTGTATTTAGTAATGGTGATTTTTTAAATT	
	>spacer6	
	CATTCATTTTTTATATATCAATAAAACTTT	
	>spacer7	

	GGGGATTCTTATTTCACTGTAGTTACGATG	
	>spacer8	
	CAAAAATTGATGTCACAATTAATAAAGGTG	
	>spacer9	
	CTATTTCTGACAATGGTTGAAATTGTGTTC	
	>spacer10	
	CTTTTTTTAAATTAATTTATCGTAAGCAA	*Streptococcus* phages: V2 / R1 / M19 / C0 / 31B4 / 16B8 / 9B4 / 7A5 / A0 / 7T / vB_SthS_VA214 / P9901 / P7601 / P7132 / P5641 / P3684 / P3681 / P0091 / 128 / 53 / 73
	>spacer11	
	AACAAACTTATGAGAACGGTTGAACGGCTT	*Streptococcus thermophilus* bacteriophage Sfi21
	>spacer12	
	AGCCCGCTTATTGCTTCAGTTGGTTTATAT	
	>spacer13	
	TGGAGCAACAAGAATGATTAACTCTAATGC	*Streptococcus* phages: P9903 / P9902 / 73
	>spacer14	
	TTTGATGGATATCATTGATAAACTATACGA	*Streptococcus* phages: STP1 / 16B8 / 7A5 / P7955 / P7954; *Streptococcus thermophilus* bacteriophage DT1
	>spacer15	
	TAACGAAAGCAATACCAATCGTGCTAAAGC	
	>spacer16	
	TATTCCTATGGTCGATATTCGAACAGTCAA	
	>spacer17	
	CAGGGGACAAGGACTTTGACCCAACAGAAG	
	>spacer18	
	AGAAACACCTAATGGTCTCTTAGAACCCGA	
	>spacer19	
	AAGAAGTTAAAGACAACTTTGTTAAAGACT	*Streptococcus* phage vB_SthS_VA214
	>spacer20	
	GAAAAAGCATCCATGATAGTGCTTAGACCT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer21	
	CGGAATGGTATAAAGAATACAAAGAAAACG	
	>spacer22	

	CCAAGTATCACGCAAAGAAATCAACGAGA	
	>spacer23	
	TTGACCTGTTTATCCTTGTTAACTAGAATAG	
	>spacer24	
	AGAGCACTAGCATACTGTTTAGTCCGAACG	
	>spacer25	
	AGGCAAGGTATTTGATCCAACAGAAGCCAA	
	>spacer26	
	CATGATTTACAACCACGCGCTAGACCAAG	
	>spacer27	
	ACCTAGAAGCATTTGAGCGTATATTGATTG	
	>spacer28	
	AATTTTGCCCCTTCTTTGCCCCTTGACTAG	*Streptococcus* phage 5093
	>spacer29	
	TAATAGTTTACCAAATCGTCCTTGTTCCAA	*Streptococcus* phage P0095
	>spacer30	
	ACCATTAGCAATCATTTGTGCCCATTGAGT	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
MN-BM-A01 (443838..445850)	>spacer1	
	GATAGCAATAGCTTTCTTGACCTAAAAGAC	*Streptococcus thermophilus* bacteriophage Sfi21
	>spacer2	
	GAGGTCTGTAATTTCATTCCCTCGTAATCT	*Streptococcus suis* 6936 plasmid unnamed1
	>spacer3	
	AAAGGTTTCTCTAAACACATGCGGAATAT	
	>spacer4	
	GTCATAGTACCAAGCACAAATAACGTTAGT	
	>spacer5	
	GTGTATTTAGTAATGGTGATTTTTTAAATT	
	>spacer6	
	CATTCATTTTTTATATATCAATAAAACTTT	
	>spacer7	
	GGGGATTCTTATTTCACTGTAGTTACGATG	
	>spacer8	
	CAAAAATTGATGTCACAATTAATAAAGGTG	
	>spacer9	

	CTATTTCTGACAATGGTTGAAATTGTGTTC	
	>spacer10	
	CTTTTTTTAAATTAATTTATCGTAAGCAA	*Streptococcus* phages: V2 / R1 / M19 / C0 / 31B4 / 16B8 / 9B4 / 7A5 / A0 / 7T / vB_SthS_VA214 / P9901 / P7601 / P7132 / P5641 / P3684 / P3681 / P0091 / 128 / 53 / 73
	>spacer11	
	AACAAACTTATGAGAACGGTTGAACGGCTT	*Streptococcus thermophilus* bacteriophage Sfi21
	>spacer12	
	AGCCCGCTTATTGCTTCAGTTGGTTTATAT	
	>spacer13	
	TGGAGCAACAAGAATGATTAACTCTAATGC	*Streptococcus* phages: P9903 / P9902 / 73
	>spacer14	
	TTTGATGGATATCATTGATAAACTATACGA	*Streptococcus* phages: STP1 / 16B8 / 7A5 / P7955 / P7954; *Streptococcus thermophilus* bacteriophage DT1
	>spacer15	
	TAACGAAAGCAATACCAATCGTGCTAAAGC	
	>spacer16	
	TATTCCTATGGTCGATATTCGAACAGTCAA	
	>spacer17	
	CAGGGGACAAGGACTTTGACCCAACAGAAG	
	>spacer18	
	AGAAACACCTAATGGTCTCTTAGAACCCGA	
	>spacer19	
	AAGAAGTTAAAGACAACTTTGTTAAAGACT	*Streptococcus* phage vB_SthS_VA214
	>spacer20	
	GAAAAAGCATCCATGATAGTGCTTAGACCT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer21	
	CGGAATGGTATAAAGAATACAAAGAAAACG	
	>spacer22	
	CCAAGTATCACGCAAAGAAATCAACGAGA	
	>spacer23	
	TTGACCTGTTTATCCTTGTTAACTAGAATAG	
	>spacer24	

	AGAGCACTAGCATACTGTTTAGTCCGAACG	
	>spacer25	
	AGGCAAGGTATTTGATCCAACAGAAGCCAA	
	>spacer26	
	CATGATTTACAACCACGCGCTAGACCAAG	
	>spacer27	
	ACCTAGAAGCATTTGAGCGTATATTGATTG	
	>spacer28	
	AATTTTGCCCCTTCTTTGCCCCTTGACTAG	*Streptococcus* phage 5093
	>spacer29	
	TAATAGTTTACCAAATCGTCCTTGTTCCAA	*Streptococcus* phage P0095
	>spacer30	
	ACCATTAGCAATCATTTGTGCCCATTGAGT	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
ASCC 1275 (823246..825392)	>spacer1	
	TGTTTGACAGCAAATCAAGATTCGAATTGT	
	>spacer2	
	AATGACGAGGAGCTATTGGCACAACTTACA	*Streptococcus* phages: M19 / L5A1 / C0 / 7A5 / 7T / vB_SthS_VA698 / P9902 / P9901 / P7955 / P7954 / P7952 / P7602 / P7134 / P3681 / Abc2 / ALQ13.2
	>spacer3	
	CGATTTGACAATCTGCTGACCACTGTTATC	
	>spacer4	
	ACACTTGGCAGGCTTATTACTCAACAGCGA	
	>spacer5	
	CTGTTCCTTGTTCTTTTGTTGTATCTTTTC	
	>spacer6	
	TTCATTCTTCCGTTTTTGTTTGCGAATCCT	
	>spacer7	
	GCTGGCGAGGAAACGAACAAGGCCTCAACA	*Streptococcus* phages: MM25 / P7152
	>spacer8	
	CATAGAGTGGAAAACTAGAAACAGATTCAA	*Streptococcus* phages: P7572 / P7571 / P7132 / 53; *Streptococcus thermophilus* bacteriophage DT1
	>spacer9	
	ATAATGCCGTTGAATTACACGGCAAGGTCA	

	>spacer10	
	GAGCGAGCTCGAAATAATCTTAATTACAAG	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952; *Streptococcus thermophilus* temperate bacteriophage O1205
	>spacer11	
	GTTCGCTAGCGTCATGTGGTAACGTATTTA	
	>spacer12	
	GGCGTCCCAATCCTGATTAATACTTACTCG	*Streptococcus* phages: P7632 / P7631 / CHPC926
	>spacer13	
	AACACAGCAAGACAAGAGGATGATGCTATG	*Streptococcus* phages: M19 / P0094 / P0093 / P0092 / P0091 / CHPC1151 / TP-J34 / 5093
	>spacer14	
	CGACACAAGAACGTATGCAAGAGTTCAAG	*Streptococcus* phages: vB_SthS_VA698 / V2 / MM25 / 31B4 / 9A / P8922 / P8921 / P7954 / P7953 / P7952 / P7632 / P7631 / P7601 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P3681 / 128 / 53 / CHPC577; *Streptococcus thermophilus* bacteriophage kappa3 / DT1
	>spacer15	
	ACAATTCTTCATCCGGTAACTGCTCAAGTG	*Streptococcus* phage vB_SthS_VA460
	>spacer16	
	AATTAAGGGCATAGAAAGGGAGACAACATG	
	>spacer17	
	CGATATTTAAAATCATTTTCATAACTTCAT	*Streptococcus* phage M19
	>spacer18	
	GCAGTATCAGCAAGCAAGCTGTTAGTTACT	*Streptococcus* phages: P9851 / YMC-2011; *Streptococcus thermophilus* bacteriophage Sfi21
	>spacer19	
	ATAAACTATGAAATTTTATAATTTTTAAGA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer20	
	AATAATTTATGGTATAGCTTAATATCATTG	*Streptococcus* phage P5641
	>spacer21	
	TGCATCGAGCACGTTCGAGTTTACCGTTTC	*Streptococcus* phage P7951
	>spacer22	

	TCTATATCGAGGTCAACTAACAATTATGCT	*Streptococcus* phage P7951
	>spacer23	
	AATCGTTCAAATTCTGTTTTAGGTACATTT	
	>spacer24	
	AATCAATACGACAAGAGTTAAAATGGTCTT	*Streptococcus* phage 5093
	>spacer25	
	GCTTAGCTGTCCAATCCACGAACGTGGATG	*Streptococcus* phage P7573; *Streptococcus thermophilus* bacteriophage 7201
	>spacer26	
	CAACCAACGGTAACAGCTACTTTTTACAGT	
	>spacer27	
	ATAACTGAAGGATAGGAGCTTGTAAAGTCT	
	>spacer28	
	TAATGCTACATCTCAAAGGATGATCCCAGA	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer29	
	AAGTAGTTGATGACCTCTACAATGGTTTAT	
	>spacer30	
	ACCTAGAAGCATTTGAGCGTATATTGATTG	
	>spacer31	
	AATTTTGCCCCTTCTTTGCCCCTTGACTAG	*Streptococcus* phage 5093
	>spacer32	
	ACCATTAGCAATCATTTGTGCCCATTGAGT	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
MN-BM-A02 (633203..635349)	>spacer1	
	TGTTTGACAGCAAATCAAGATTCGAATTGT	
	>spacer2	
	AATGACGAGGAGCTATTGGCACAACTTACA	*Streptococcus* phages: M19 / L5A1 / C0 / 7A5 / 7T / vB_SthS_VA698 / P9902 / P9901 / P7955 / P7954 / P7952 / P7602 / P7134 / P3681 / Abc2 / ALQ13.2
	>spacer3	
	CGATTTGACAATCTGCTGACCACTGTTATC	
	>spacer4	
	ACACTTGGCAGGCTTATTACTCAACAGCGA	
	>spacer5	
	CTGTTCCTTGTTCTTTTGTTGTATCTTTTC	

	>spacer6	
	TTCATTCTTCCGTTTTTGTTTGCGAATCCT	
	>spacer7	
	GCTGGCGAGGAAACGAACAAGGCCTCAACA	*Streptococcus* phages: MM25 / P7152
	>spacer8	
	CATAGAGTGGAAAACTAGAAACAGATTCAA	*Streptococcus* phages: P7572 / P7571 / P7132 / 53; *Streptococcus thermophilus* bacteriophage DT1
	>spacer9	
	ATAATGCCGTTGAATTACACGGCAAGGTCA	
	>spacer10	
	GAGCGAGCTCGAAATAATCTTAATTACAAG	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952; *Streptococcus thermophilus* temperate bacteriophage O1205
	>spacer11	
	GTTCGCTAGCGTCATGTGGTAACGTATTTA	
	>spacer12	
	GGCGTCCCAATCCTGATTAATACTTACTCG	*Streptococcus* phages: P7632 / P7631 / CHPC926
	>spacer13	
	AACACAGCAAGACAAGAGGATGATGCTATG	*Streptococcus* phages: M19 / P0094 / P0093 / P0092 / P0091 / CHPC1151 / TP-J34 / 5093
	>spacer14	
	CGACACAAGAACGTATGCAAGAGTTCAAG	*Streptococcus* phages: vB_SthS_VA698 / V2 / MM25 / 31B4 / 9A / P8922 / P8921 / P7954 / P7953 / P7952 / P7632 / P7631 / P7601 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P3681 / 128 / 53 / CHPC577; *Streptococcus thermophilus* bacteriophage kappa3 / DT1
	>spacer15	
	ACAATTCTTCATCCGGTAACTGCTCAAGTG	*Streptococcus* phage vB_SthS_VA460
	>spacer16	
	AATTAAGGGCATAGAAAGGGAGACAACATG	
	>spacer17	
	CGATATTTAAAATCATTTTCATAACTTCAT	*Streptococcus* phage M19
	>spacer18	

	GCAGTATCAGCAAGCAAGCTGTTAGTTACT	*Streptococcus* phages: P9851 / YMC-2011; *Streptococcus thermophilus* bacteriophage Sfi21
	>spacer19	
	ATAAACTATGAAATTTTATAATTTTTAAGA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer20	
	AATAATTTATGGTATAGCTTAATATCATTG	*Streptococcus* phage P5641
	>spacer21	
	TGCATCGAGCACGTTCGAGTTTACCGTTTC	*Streptococcus* phage P7951
	>spacer22	
	TCTATATCGAGGTCAACTAACAATTATGCT	*Streptococcus* phage P7951
	>spacer23	
	AATCGTTCAAATTCTGTTTTAGGTACATTT	
	>spacer24	
	AATCAATACGACAAGAGTTAAAATGGTCTT	*Streptococcus* phage 5093
	>spacer25	
	GCTTAGCTGTCCAATCCACGAACGTGGATG	*Streptococcus* phage P7573; *Streptococcus thermophilus* bacteriophage 7201
	>spacer26	
	CAACCAACGGTAACAGCTACTTTTTACAGT	
	>spacer27	
	ATAACTGAAGGATAGGAGCTTGTAAAGTCT	
	>spacer28	
	TAATGCTACATCTCAAAGGATGATCCCAGA	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer29	
	AAGTAGTTGATGACCTCTACAATGGTTTAT	
	>spacer30	
	ACCTAGAAGCATTTGAGCGTATATTGATTG	
	>spacer31	
	AATTTTGCCCCTTCTTTGCCCCTTGACTAG	*Streptococcus* phage 5093
	>spacer32	
	ACCATTAGCAATCATTTGTGCCCATTGAGT	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
KLDS SM (1806995..1809141)	>spacer1	
	TGTTTGACAGCAAATCAAGATTCGAATTGT	

	>spacer2	
	AATGACGAGGAGCTATTGGCACAACTTACA	*Streptococcus* phages: M19 / L5A1 / C0 / 7A5 / 7T / vB_SthS_VA698 / P9902 / P9901 / P7955 / P7954 / P7952 / P7602 / P7134 / P3681 / Abc2 / ALQ13.2
	>spacer3	
	CGATTTGACAATCTGCTGACCACTGTTATC	
	>spacer4	
	ACACTTGGCAGGCTTATTACTCAACAGCGA	
	>spacer5	
	CTGTTCCTTGTTCTTTTGTTGTATCTTTTC	
	>spacer6	
	TTCATTCTTCCGTTTTTGTTTGCGAATCCT	
	>spacer7	
	GCTGGCGAGGAAACGAACAAGGCCTCAACA	*Streptococcus* phages: MM25 / P7152
	>spacer8	
	CATAGAGTGGAAAACTAGAAACAGATTCAA	*Streptococcus* phages: P7572 / P7571 / P7132 / 53; *Streptococcus thermophilus* bacteriophage DT1
	>spacer9	
	ATAATGCCGTTGAATTACACGGCAAGGTCA	
	>spacer10	
	GAGCGAGCTCGAAATAATCTTAATTACAAG	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952; *Streptococcus thermophilus* temperate bacteriophage O1205
	>spacer11	
	GTTCGCTAGCGTCATGTGGTAACGTATTTA	
	>spacer12	
	GGCGTCCCAATCCTGATTAATACTTACTCG	*Streptococcus* phages: P7632 / P7631 / CHPC926
	>spacer13	
	AACACAGCAAGACAAGAGGATGATGCTATG	*Streptococcus* phages: M19 / P0094 / P0093 / P0092 / P0091 / CHPC1151 / TP-J34 / 5093
	>spacer14	

	CGACACAAGAACGTATGCAAGAGTTCAAG	*Streptococcus* phages: vB_SthS_VA698 / V2 / MM25 / 31B4 / 9A / P8922 / P8921 / P7954 / P7953 / P7952 / P7632 / P7631 / P7601 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P3681 / 128 / 53 / CHPC577; *Streptococcus thermophilus* bacteriophage kappa3 / DT1
	>spacer15	
	ACAATTCTTCATCCGGTAACTGCTCAAGTG	*Streptococcus* phage vB_SthS_VA460
	>spacer16	
	AATTAAGGGCATAGAAAGGGAGACAACATG	
	>spacer17	
	CGATATTTAAAATCATTTTCATAACTTCAT	*Streptococcus* phage M19
	>spacer18	
	GCAGTATCAGCAAGCAAGCTGTTAGTTACT	*Streptococcus* phages: P9851 / YMC-2011; *Streptococcus thermophilus* bacteriophage Sfi21
	>spacer19	
	ATAAACTATGAAATTTTATAATTTTTAAGA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer20	
	AATAATTTATGGTATAGCTTAATATCATTG	*Streptococcus* phage P5641
	>spacer21	
	TGCATCGAGCACGTTCGAGTTTACCGTTTC	*Streptococcus* phage P7951
	>spacer22	
	TCTATATCGAGGTCAACTAACAATTATGCT	*Streptococcus* phage P7951
	>spacer23	
	AATCGTTCAAATTCTGTTTTAGGTACATTT	
	>spacer24	
	AATCAATACGACAAGAGTTAAAATGGTCTT	*Streptococcus* phage 5093
	>spacer25	
	GCTTAGCTGTCCAATCCACGAACGTGGATG	*Streptococcus* phage P7573; *Streptococcus thermophilus* bacteriophage 7201
	>spacer26	
	CAACCAACGGTAACAGCTACTTTTTACAGT	
	>spacer27	
	ATAACTGAAGGATAGGAGCTTGTAAAGTCT	

	>spacer28	
	TAATGCTACATCTCAAAGGATGATCCCAGA	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer29	
	AAGTAGTTGATGACCTCTACAATGGTTTAT	
	>spacer30	
	ACCTAGAAGCATTTGAGCGTATATTGATTG	
	>spacer31	
	AATTTTGCCCCTTCTTTGCCCCTTGACTAG	*Streptococcus* phage 5093
	>spacer32	
	ACCATTAGCAATCATTTGTGCCCATTGAGT	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
ND07 (177242..179388) (rev_com)	>spacer1	
	TGTTTGACAGCAAATCAAGATTCGAATTGT	
	>spacer2	
	AATGACGAGGAGCTATTGGCACAACTTACA	*Streptococcus* phages: M19 / L5A1 / C0 / 7A5 / 7T / vB_SthS_VA698 / P9902 / P9901 / P7955 / P7954 / P7952 / P7602 / P7134 / P3681 / Abc2 / ALQ13.2
	>spacer3	
	CGATTTGACAATCTGCTGACCACTGTTATC	
	>spacer4	
	ACACTTGGCAGGCTTATTACTCAACAGCGA	
	>spacer5	
	CTGTTCCTTGTTCTTTTGTTGTATCTTTTC	
	>spacer6	
	TTCATTCTTCCGTTTTTGTTTGCGAATCCT	
	>spacer7	
	GCTGGCGAGGAAACGAACAAGGCCTCAACA	*Streptococcus* phages: MM25 / P7152
	>spacer8	
	CATAGAGTGGAAAACTAGAAACAGATTCAA	*Streptococcus* phages: P7572 / P7571 / P7132 / 53; *Streptococcus thermophilus* bacteriophage DT1
	>spacer9	
	ATAATGCCGTTGAATTACACGGCAAGGTCA	
	>spacer10	

	GAGCGAGCTCGAAATAATCTTAATTACAAG	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952; *Streptococcus thermophilus* temperate bacteriophage O1205
	>spacer11	
	GTTCGCTAGCGTCATGTGGTAACGTATTTA	
	>spacer12	
	GGCGTCCCAATCCTGATTAATACTTACTCG	*Streptococcus* phages: P7632 / P7631 / CHPC926
	>spacer13	
	AACACAGCAAGACAAGAGGATGATGCTATG	*Streptococcus* phages: M19 / P0094 / P0093 / P0092 / P0091 / CHPC1151 / TP-J34 / 5093
	>spacer14	
	CGACACAAGAACGTATGCAAGAGTTCAAG	*Streptococcus* phages: vB_SthS_VA698 / V2 / MM25 / 31B4 / 9A / P8922 / P8921 / P7954 / P7953 / P7952 / P7632 / P7631 / P7601 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P3681 / 128 / 53 / CHPC577; *Streptococcus thermophilus* bacteriophage kappa3 / DT1
	>spacer15	
	ACAATTCTTCATCCGGTAACTGCTCAAGTG	*Streptococcus* phage vB_SthS_VA460
	>spacer16	
	AATTAAGGGCATAGAAAGGGAGACAACATG	
	>spacer17	
	CGATATTTAAAATCATTTTCATAACTTCAT	*Streptococcus* phage M19
	>spacer18	
	GCAGTATCAGCAAGCAAGCTGTTAGTTACT	*Streptococcus* phages: P9851 / YMC-2011; *Streptococcus thermophilus* bacteriophage Sfi21
	>spacer19	
	ATAAACTATGAAATTTTATAATTTTTAAGA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer20	
	AATAATTTATGGTATAGCTTAATATCATTG	*Streptococcus* phage P5641
	>spacer21	
	TGCATCGAGCACGTTCGAGTTTACCGTTTC	*Streptococcus* phage P7951
	>spacer22	
	TCTATATCGAGGTCAACTAACAATTATGCT	*Streptococcus* phage P7951

	>spacer23	
	AATCGTTCAAATTCTGTTTTAGGTACATTT	
	>spacer24	
	AATCAATACGACAAGAGTTAAAATGGTCTT	*Streptococcus* phage 5093
	>spacer25	
	GCTTAGCTGTCCAATCCACGAACGTGGATG	*Streptococcus* phage P7573; *Streptococcus thermophilus* bacteriophage 7201
	>spacer26	
	CAACCAACGGTAACAGCTACTTTTTACAGT	
	>spacer27	
	ATAACTGAAGGATAGGAGCTTGTAAAGTCT	
	>spacer28	
	TAATGCTACATCTCAAAGGATGATCCCAGA	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer29	
	AAGTAGTTGATGACCTCTACAATGGTTTAT	
	>spacer30	
	ACCTAGAAGCATTTGAGCGTATATTGATTG	
	>spacer31	
	AATTTTGCCCCTTCTTTGCCCCTTGACTAG	*Streptococcus* phage 5093
	>spacer32	
	ACCATTAGCAATCATTTGTGCCCATTGAGT	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
DGCC 7710 (636570..638716)	>spacer1	
	TGTTTGACAGCAAATCAAGATTCGAATTGT	
	>spacer2	
	AATGACGAGGAGCTATTGGCACAACTTACA	*Streptococcus* phages: M19 / L5A1 / C0 / 7A5 / 7T / vB_SthS_VA698 / P9902 / P9901 / P7955 / P7954 / P7952 / P7602 / P7134 / P3681 / Abc2 / ALQ13.2
	>spacer3	
	CGATTTGACAATCTGCTGACCACTGTTATC	
	>spacer4	
	ACACTTGGCAGGCTTATTACTCAACAGCGA	
	>spacer5	
	CTGTTCCTTGTTCTTTTGTTGTATCTTTTC	
	>spacer6	

	TTCATTCTTCCGTTTTTGTTTGCGAATCCT	
	>spacer7	
	GCTGGCGAGGAAACGAACAAGGCCTCAACA	*Streptococcus* phages: MM25 / P7152
	>spacer8	
	CATAGAGTGGAAAACTAGAAACAGATTCAA	*Streptococcus* phages: P7572 / P7571 / P7132 / 53; *Streptococcus thermophilus* bacteriophage DT1
	>spacer9	
	ATAATGCCGTTGAATTACACGGCAAGGTCA	
	>spacer10	
	GAGCGAGCTCGAAATAATCTTAATTACAAG	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952; *Streptococcus thermophilus* temperate bacteriophage O1205
	>spacer11	
	GTTCGCTAGCGTCATGTGGTAACGTATTTA	
	>spacer12	
	GGCGTCCCAATCCTGATTAATACTTACTCG	*Streptococcus* phages: P7632 / P7631 / CHPC926
	>spacer13	
	AACACAGCAAGACAAGAGGATGATGCTATG	*Streptococcus* phages: M19 / P0094 / P0093 / P0092 / P0091 / CHPC1151 / TP-J34 / 5093
	>spacer14	
	CGACACAAGAACGTATGCAAGAGTTCAAG	*Streptococcus* phages: vB_SthS_VA698 / V2 / MM25 / 31B4 / 9A / P8922 / P8921 / P7954 / P7953 / P7952 / P7632 / P7631 / P7601 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P3681 / 128 / 53 / CHPC577; *Streptococcus thermophilus* bacteriophage kappa3 / DT1
	>spacer15	
	ACAATTCTTCATCCGGTAACTGCTCAAGTG	*Streptococcus* phage vB_SthS_VA460
	>spacer16	
	AATTAAGGGCATAGAAAGGGAGACAACATG	
	>spacer17	
	CGATATTTAAAATCATTTTCATAACTTCAT	*Streptococcus* phage M19
	>spacer18	
	GCAGTATCAGCAAGCAAGCTGTTAGTTACT	*Streptococcus* phages: P9851 / YMC-2011; *Streptococcus thermophilus* bacteriophage Sfi21

	>spacer19	
	ATAAACTATGAAATTTTATAATTTTTAAGA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer20	
	AATAATTTATGGTATAGCTTAATATCATTG	*Streptococcus* phage P5641
	>spacer21	
	TGCATCGAGCACGTTCGAGTTTACCGTTTC	*Streptococcus* phage P7951
	>spacer22	
	TCTATATCGAGGTCAACTAACAATTATGCT	*Streptococcus* phage P7951
	>spacer23	
	AATCGTTCAAATTCTGTTTTAGGTACATTT	
	>spacer24	
	AATCAATACGACAAGAGTTAAAATGGTCTT	*Streptococcus* phage 5093
	>spacer25	
	GCTTAGCTGTCCAATCCACGAACGTGGATG	*Streptococcus* phage P7573; *Streptococcus thermophilus* bacteriophage 7201
	>spacer26	
	CAACCAACGGTAACAGCTACTTTTTACAGT	
	>spacer27	
	ATAACTGAAGGATAGGAGCTTGTAAAGTCT	
	>spacer28	
	TAATGCTACATCTCAAAGGATGATCCCAGA	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer29	
	AAGTAGTTGATGACCTCTACAATGGTTTAT	
	>spacer30	
	ACCTAGAAGCATTTGAGCGTATATTGATTG	
	>spacer31	
	AATTTTGCCCCTTCTTTGCCCCTTGACTAG	*Streptococcus* phage 5093
	>spacer32	
	ACCATTAGCAATCATTTGTGCCCATTGAGT	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
KLDS 3.1003 (1254932..1255890)	>spacer1	
	CGATTGGATCACCAAAGCTTTGTCTGATG	*Streptococcus* phages: P7954 / P7953 / P7952
	>spacer2	
	TCAAAAAACTGTGATGAGACTTTAGGTTGG	

	>spacer3	
	CACTTTCAAGCGATCGAGTCCATTTGTCGT	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952
	>spacer4	
	CTAAAATTGATGTAACAATCAATAAAGGTG	
	>spacer5	
	CTGTCTGAAAAATAACCGACCATCGTTCGA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer6	
	GTCAGTACCATCAATATGCTGGACGGCACG	*Streptococcus thermophilus* bacteriophages: 7201 / Sfi19
	>spacer7	
	ATGGTAGCTCCGCTCTTGGCTTGTGCTTCA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer8	
	TGAGGAATTATTCAGACAGTTTTTAAACTA	*Streptococcus* phage TP-J34
	>spacer9	
	CCTTCCTAAGTGCATGAAAATCGCAAACGG	*Streptococcus thermophilus* bacteriophages: Sfi11 / 7201 / Sfi21 / Sfi19
	>spacer10	
	CAAAGTCCTCTTCGTATTGATCATAGCTTC	*Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* FM03 plasmid pLd7 / *Lactococcus lactis* subsp. *cremoris* JM2 plasmid pJM2C / *Lactococcus lactis* subsp. *lactis* 229 plasmid p229C / *Lactococcus garvieae* IPLA 31405 plasmid pLG42 / *Lactococcus lactis* subsp. *lactis* ILIBB-JZK plasmid pIBB-JZK / *Lactococcus lactis* subsp. *lactis* bv. diacetylactis plasmid pVF21 / *Lactococcus lactis* subsp. *lactis* IL594 plasmid pIL6 / *Lactococcus lactis* subsp. *lactis* CV56 plasmid pCV56B / *Lactococcus lactis* plasmid pIL2614 / *Lactococcus lactis* subsp. *lactis* plasmid pAH82 / *Lactococcus lactis* plasmid pAH82 / *Lactococcus lactis* subsp. *lactis* strain UC063 plasmid pUC063B
	>spacer11	
	AGGTCGAAGTTTGGGAAATTGACTTGGCTG	*Streptococcus* phage TP-J34; *Streptococcus thermophilus* bacteriophage Sfi11 / O1205
	>spacer12	

	TTGATAAGTCTTATATTGTCACAGGAAAGA	
	>spacer13	
	TAGACATCTACGACTTATTTTTAAAGGCAG	
	>spacer14	
	GACGTGTCAACACGGTTAGTTAACATTGAT	
B59671 (1815698..1816920)	>spacer1	
	AAATTGATATCTATACATCAAGCTGGATAA	*Streptococcus* phages: P9903 / P9902 / P7602 / 73
	>spacer2	
	CTGAATAGTATTTCTTGATTGGCTCGTTGC	*Streptococcus thermophilus* bacteriophages: 7201 / Sfi19
	>spacer3	
	TGATTTTCTCAAACCAGAGCGCAGCACCCA	
	>spacer4	
	TAGACAAAAACCTTCTATCTCCCAAAACTG	
	>spacer5	
	TGGCTAAGATAGGGTATGCGCGTGTAAGTA	
	>spacer6	
	GTGTTCCAGCGTCTTTGTATTTATTTCGGA	*Streptococcus* phages: L5A1 / C0 / A0
	>spacer7	
	CTTTGTTTCCGCCTTTGACCACGTCGGAAC	
	>spacer8	
	TAATCACCCCTTTTTCTAGCTCTTGATTGA	
	>spacer9	
	AAGTTGATCGTATCTATTTAGAATATCGCA	*Streptococcus* phages: P9852 / P7955 / P7953 / P7951 / P7574 / P7572 / P7571 / TP-J34 / 858; *Streptococcus thermophilus* bacteriophages: 2972 / Sfi11 / Sfi21 / Sfi19
	>spacer10	
	TGCTATGCAAATAGCAGGAGCGATGATTG	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952 / P4761 / TP-778L / TP-J34; *Streptococcus thermophilus* bacteriophages: Sfi11 / O1205
	>spacer11	
	CATATCTGACAGAAACAGTATCGTTGGCTT	*Lactococcus* phage 712
	>spacer12	

	AATCTTACGACCCTAACATTTTCTACCAAC	
	>spacer13	
	AGATAGCCACCAAGCGTTTCAGCTATGCTA	*Streptococcus thermophilus* bacteriophage Sfi19
	>spacer14	
	TACTTGAACACCCTTCAATCCTGTTTGATA	*Streptococcus* phages: L5A1 / C0 / A0 / P9854 / P9851 / P7573 / vB_SthS_VA460 / P7602 / P5641 / P3684 / P3681
	>spacer15	
	TTTTCAGCTTCAGTCGTTTCATATGTATAT	
	>spacer16	
	AAGGTCTATACCCTAATATAGCTGAACACG	
	>spacer17	
	ACATGCTTGATGTTTGTGTATTTAGTAACA	
	>spacer18	
	GAGTTGGTGACCATTTGCACCACTTTATTT	
NCTC12958 (803250..805592)	>spacer1	
	ATATACGGACTCATCTAGTGAAACTGGGTA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer2	
	GAATCTGTTAAAACATTGATTTTTTATATT	
	>spacer3	
	TGTGCCAGCGTCATTCAAAGCAATTGCTGA	*Streptococcus* phages: 9A / P3684 / P3681 / P0091; *Streptococcus thermophilus* bacteriophage Sfi21
	>spacer4	
	CAAAGCTATCGAATATATCGCAGAGGCAAG	*Streptococcus* phages: P0095 / TP-J34 / 5093 / P7602; *Streptococcus thermophilus* bacteriophage Sfi19
	>spacer5	
	TCTTTGTGATATAAATGAAAAAAAAGACTA	
	>spacer6	
	CAAGCGTAAGTTTTACCGTTAGAACGTTCA	
	>spacer7	
	GATAGCGCCTTTGAATATTAATGGTGTTC	*Streptococcus* phages: vB_SthS_VA214 / P7955 / P7953 / P7952 / P7951 / P0091 / Abc2
	>spacer8	

	TAGACTTTACTAATTCTCAAGTAAATATAT	*Streptococcus* phages: P9903 / P9902 / 73 / TP-778L / TP-J34
	>spacer9	
	TTCTTTCAGGTATGAAATTAGAAACACTTG	
	>spacer10	
	TTTCACAGTCAACACATCTCCGACACGCTT	*Streptococcus* phages: P7602 / M19 / vB_SthS_VA698 / ALQ13.2
	>spacer11	
	ATGAGTTGGTAGATGTTAAAGCATCAACTG	
	>spacer12	
	GTATATACAGAACTTGGAGTTTTAACGCCT	*Streptococcus* phages: P7954 / P7953 / P7952 / P3684 / CHPC926
	>spacer13	
	CATACACCACCAGCAAAGTCTTTACCAATG	
	>spacer14	
	GATGATTTTGAGTTTAAGGTTTTTAAACAA	*Streptococcus* phage D4276
	>spacer15	
	AGACCATGAAGGAAGGCGTTGACAAATTC	*Streptococcus* phages: P7602 / P3684 / P3681
	>spacer16	
	AGAAATTCATGATAAGTGGTCTGAAACTTG	*Streptococcus* phage vB_SthS_VA460
	>spacer17	
	AGTCAACTTACTATGTTAAAATAAGACAGG	
	>spacer18	
	GACTTGAACCTGAATTAATCAAGGTAGCCA	*Streptococcus* phages: STP2 / P0094 / P0093 / P0091 / CHPC1151 / 9871
	>spacer19	
	TGTACTCTATTGATTGCTTCATCTTTATTA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer20	
	CTTTCAAGATACTCATCAACCATTGATGTCA	
	>spacer21	
	AAAAGGTGCGTATGAAACTCATCCCAGCGG	
	>spacer22	
	CTATGTCTTTACTGTTCTTCCAAAACCACC	*Streptococcus* phages: P0094 / P0093 / 5093
	>spacer23	

	TTTGTTTGCGTTCACTTAGTTTATACTGAG	*Streptococcus* phages: D1811 / D1024 / STP2 / STP1 / vB_SthS_VA460 / vB_SthS_VA214 / P9854 / P9853 / P9852 / P9851 / P8921 / P7955 / P7573 / P7572 / P7571 / P7152 / P7151 / P7134 / P5652 / P5651 / P4761 / P0095 / P0094 / P0093 / P0092 / CHPC1151 / CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871 / ALQ13.2 / 858; *Streptococcus thermophilus* bacteriophages: 2972 / DT1 / J1 / ST3 / S92 / S3b
	>spacer24	
	AATTTGTCATTGACATTTACCACCGTCCTG	Streptococcus phages: P9854 / P9851 / D4276
	>spacer25	
	TGGCGCAATGGTTTTAAATCATATGCCTT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer26	
	CATGGCCTTGTATTTTCGAAAGAGCATAAG	*Streptococcus* phage P4761
	>spacer27	
	TATCCTCTCGTTGATAATATGCTTAATTCT	
	>spacer28	
	TTGATAGACCAAAACACTTCTGTGAGCAGC	
	>spacer29	
	TGCCTTTCTTCAAAGAAGCTATGCTTGTTA	*Streptococcus* phages: P7955 / P7953 / P7952
	>spacer30	
	CTGCATCAAACGATGGCTTTGTCATCGGT	
	>spacer31	
	AATATAATCGACGAGTAAGAGCGAGAGACG	
	>spacer32	
	CGCGCACCGTTGACAAGTGTCAAGTTATCG	*Streptococcus* phage ALQ13.2
	>spacer33	
	GATGATTTTGAGTTTAAGGTTTTTAAACAA	*Streptococcus* phage D4276
	>spacer34	
	TCATACTAAAACTGGATAAACAATATTACA	
	>spacer35	
	TGACACGGCTGGCAGTGTTTCGGTTGGTGG	*Streptococcus* phages: M19 / P9903 / P9902 / ALQ13.2; *Streptococcus thermophilus* bacteriophages: DT1.1 / DT1.2 / DT1.3 / DT1.4 / DT1.5 / MD4 / DT1

Spacers of CRISPR2		
Strain (coordinates of CRISPR array)	**spacers**	**1st blastn result (excluding same species)**
LMG 18311 (863999..864323)	>spacer1	
	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
	>spacer2	
	TTTGAATAGTCTTTGAATCGCATTTGAACCATATA	*Streptococcus* phages: 9A / P8922 / P8921 / P7953 / P7154 / P7151 / P5652 / P5651 / 9874 / ALQ13.2
	>spacer3	
	AGGTTTTTTGCCATAGATTTTCCAAGACCTTCCCAACT	
	>spacer4	
	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
KLDS 3.1003 (1506836..1507597)	>spacer1	
	CCTACAATAGCATTCATTGTAAATTGTTGCATTTTT	
	>spacer2	
	TCATCTGGTAACTGCTCTAGTGTTAGTATGCGATTG	*Streptococcus* phages: P7574 / vB_SthS_VA460 / P0095 / P7602 / P7601 / P7571 / P7132 / 53; *Streptococcus thermophilus* bacteriophage DT1
	>spacer3	
	TCACGGTCCTTTTTCATCTGTCTCAAAATTTCAA	*Streptococcus* phage 5093
	>spacer4	
	TGTTCCATATTGGTAGCATTAATGAATCTGTACTTATCT	*Streptococcus thermophilus* bacteriophage Sfi21
	>spacer5	
	ACCCCAACTGGTTTGTAGGTTATTTCCATTGGTAGTCC	
	>spacer6	
	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
	>spacer7	
	TGAAAGCACTATGGAATGATACATTAGTAGTTCTGG	*Streptococcus* phage 5093
	>spacer8	
	GACAGAATTGTGGACACTGCCATACTAATAGATGTACAG	*Streptococcus* phages: TP-778L / TP-J34
	>spacer9	
	AGATGTTTTCATGGTATACTCCTTTTAAGTTGTTTT	*Streptococcus thermophilus* bacteriophage O1205
	>spacer10	

	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
JIM 8232 (969817..971087)	>spacer1	
	CTTGATTCTAACGCTACTTCTAAATAAGCGTTAGCAATG	*Streptococcus* phage 5093
	>spacer2	
	AGTTCTAAAGCATTCATCATATGTTTATGTTCTTTA	*Streptococcus* phages: vB_SthS_VA214 / P0095 / P0091 / CHPC1151
	>spacer3	
	TTCCAAAATTAATTCTTTGATGTCGTTATTTATTA	*Streptococcus* virus 9874
	>spacer4	
	ATCCAGTTCTGCAAAGAACAGCATAGAGGTTTTACC	*Streptococcus* phage ALQ13.2
	>spacer5	
	TTTTAATTCTTCTAAGTTATCGCTGAATGGATAGAAC	
	>spacer6	
	ACTGTTTCTGTTTGCTTGTTTGACCAAATCCAGTTAA	*Streptococcus* phages: P7572 / P7571; *Streptococcus thermophilus* bacteriophage DT1
	>spacer7	
	AATTGGATAGCGCAGGCTACCAAACTTCTAAGATGATT	
	>spacer8	
	TCTTGTTGGCGTTTTTCAATAACTTTCTTATCAATCA	*Streptococcus* phages: vB_SthS_VA460 / D4276 / TP-J34; *Streptococcus thermophilus* bacteriophages: Sfi11 / SFi18 / Sfi19
	>spacer9	
	TTCCAAAATTAATTCTTTGATGTCGTTATTTATTA	*Streptococcus* virus 9874
	>spacer10	
	ATCCAGTTCTGCAAAGAACAGCATAGAGGTTTTACC	*Streptococcus* phage ALQ13.2
	>spacer11	
	TTTTAATTCTTCTAAGTTATCGCTGAATGGATAGAAC	
	>spacer12	
	ACTGTTTCTGTTTGCTTGTTTGACCAAATCCAGTTAA	*Streptococcus* phages: P7572 / P7571; *Streptococcus thermophilus* bacteriophage DT1
	>spacer13	
	AATTGGATAGCGCAGGCTACCAAACTTCTAAGATGATT	
	>spacer14	

	TCTTGTTGGCGTTTTTCAATAACTTTCTTATCAATCA	*Streptococcus* phages: vB_SthS_VA460 / D4276 / TP-J34; *Streptococcus thermophilus* bacteriophages: Sfi11 / SFi18 / Sfi19
	>spacer15	
	TGAGAAACATAGAAAACATGGTATGAATTATCATAG	*Streptococcus* phages: P9854 / P9851 / P7574 / P7573; *Streptococcus thermophilus* bacteriophage 7201
	>spacer16	
	CCATATACTCTAGCATTTCCATACACTTTAGCATTT	
	>spacer17	
	TTAACGAAGTTGTCATCAATCATTACTCCTTGCCTG	*Streptococcus* phages: P7954 / P7953 / P7951 / P7574 / P7573 / P7132; *Streptococcus thermophilus* bacteriophages: Sfi11 / SFi18 / Sfi19
LMD-9 (897070..897328)	>spacer1	
	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
	>spacer2	
	TCAAAATATGTTATTACCTTGTATTTCATAATTCAATTAA	*Streptococcus* phage 73
	>spacer3	
	CCACTTGCTGTGTACATCCTACCAGTTCCGCCTATGATG	*Streptococcus* phages: P9903 / P9902 / 73; *Streptococcus thermophilus* bacteriophage DT1
ASCC 1275 (1074866..1075124)	>spacer1	
	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
	>spacer2	
	TCAAAATATGTTATTACCTTGTATTTCATAATTCAATTAA	*Streptococcus* phage 73
	>spacer3	
	CCACTTGCTGTGTACATCCTACCAGTTCCGCCTATGATG	*Streptococcus* phages: P9903 / P9902 / 73; *Streptococcus thermophilus* bacteriophage DT1
SMQ-301 (897843..898101)	>spacer1	
	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
	>spacer2	
	TCAAAATATGTTATTACCTTGTATTTCATAATTCAATTAA	*Streptococcus* phage 73
	>spacer3	
	CCACTTGCTGTGTACATCCTACCAGTTCCGCCTATGATG	*Streptococcus* phages: P9903 / P9902 / 73; *Streptococcus thermophilus* bacteriophage DT1
MN-BM-A02 (884614..884872)	>spacer1	
	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	

	>spacer2	
	TCAAAATATGTTATTACCTTGTATTTCATAATTCAATTAA	*Streptococcus* phage 73
	>spacer3	
	CCACTTGCTGTGTACATCCTACCAGTTCCGCCTATGATG	*Streptococcus* phages: P9903 / P9902 / 73; *Streptococcus thermophilus* bacteriophage DT1
KLDS SM (199829..200087)	>spacer1	
	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
	>spacer2	
	TCAAAATATGTTATTACCTTGTATTTCATAATTCAATTAA	*Streptococcus* phage 73
	>spacer3	
	CCACTTGCTGTGTACATCCTACCAGTTCCGCCTATGATG	*Streptococcus* phages: P9903 / P9902 / 73; *Streptococcus thermophilus* bacteriophage DT1
ND07 (1779078..1779330) (rev_com)	>spacer1	
	GGAAACGCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
	>spacer2	
	GGAAACTCAAAATATGTTATTACCTTGTATTTCATAATTCAATTAA	*Streptococcus* phage 73
	>spacer3	
	GGAAACCCACTTGCTGTGTACATCCTACCAGTTCCGCCTATGATG	*Streptococcus* phages: P9903 / P9902 / 73; *Streptococcus thermophilus* bacteriophage DT1
DGCC 7710 (887984..888242)	>spacer1	
	GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA	
	>spacer2	
	TCAAAATATGTTATTACCTTGTATTTCATAATTCAATTAA	*Streptococcus* phage 73
	>spacer3	
	CCACTTGCTGTGTACATCCTACCAGTTCCGCCTATGATG	*Streptococcus* phages: P9903 / P9902 / 73; *Streptococcus thermophilus* bacteriophage DT1
GABA (889434..889684)	>spacer1	
	TGAGTGCTTATCAAAAGCTAGAAAAGAGTTATAAAG	*Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* FM03 plasmid pLd7 / *Lactococcus lactis* subsp. *cremoris* JM2 plasmid pJM2A / *Lactococcus lactis* subsp. *lactis* 229 plasmid p229C / *Lactococcus lactis* subsp. *lactis* IL594 plasmid pIL6 / *Lactococcus lactis* subsp. *lactis* CV56 plasmid pCV56B / *Lactococcus lactis* bv. *diacetylactis*, plasmid pND861
	>spacer2	

	CAATCCACAATCTAATGCGATACCCACGGTAATTA	*Streptococcus* phage YMC-2011
	>spacer3	
	TGTCCGCTGGATATTTATCGCTTTAGTAATTTTCAT	*Streptococcus* phages: P7602 / P7134 / P7133 / P3681
ST3 (914332..914582)	>spacer1	
	TGAGTGCTTATCAAAAGCTAGAAAAGAGTTATAAAG	*Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* FM03 plasmid pLd7 / *Lactococcus lactis* subsp. *cremoris* JM2 plasmid pJM2A / *Lactococcus lactis* subsp. *lactis* 229 plasmid p229C / *Lactococcus lactis* subsp. *lactis* IL594 plasmid pIL6 / *Lactococcus lactis* subsp. *lactis* CV56 plasmid pCV56B / *Lactococcus lactis* bv. *diacetylactis*, plasmid pND861
	>spacer2	
	CAATCCACAATCTAATGCGATACCCACGGTAATTA	*Streptococcus* phage YMC-2011
	>spacer3	
	TGTCCGCTGGATATTTATCGCTTTAGTAATTTTCAT	*Streptococcus* phages: P7602 / P7134 / P7133 / P3681
Spacers of CRISPR3		
Strain (coordinates of CRISPR array)	**spacers**	**1st blastn result (excluding same species)**
KLDS 3.1003 (121258..122019)	>spacer1	
	TGACGAAACGGAAAGTAATACAGATAACT	
	>spacer2	
	AGCTAGACATATCAGCACCCCATGACATT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer3	
	TGACCATATTATAGGTTGAGTTGGTTTTT	
	>spacer4	
	CGTGTCCACCAGCTCCACCTGAACTTGCC	*Streptococcus* phage TP-J34
	>spacer5	
	AATCAGAAATCAGTTTATTGATTCCATCG	*Streptococcus* phages: P8921 / P7152; *Streptococcus thermophilus* bacteriophage Sfi21
	>spacer6	
	CTTTACCTGGGAAACCATCATTACCTCT	
	>spacer7	
	CCACTGGTTGCGTTTGGGATGTTATCAGC	
	>spacer8	

	ACTTAAACGACATCATTTTTATAACTTCG	*Streptococcus* virus 9874
	>spacer9	
	TGATGTAAGATAATTTAGTTCCTAATTGT	
	>spacer10	
	CACACTTATTCCGTGATGAAGTCGGTTGG	*Streptococcus* phages: P0095 / P0094 / P0093 / P0092 / CHPC1151 / 5093
	>spacer11	
	AGCTTTTTGGAATATCGAAATTAACCGTT	
ND03 (1363057..1364415)	>spacer1	
	ATAAGATTGAAACATACTTGTGCAAGTGTG	*Streptococcus* phage D4276
	>spacer2	
	AATTATCTTGGTAGAAAACTTTATTATTTA	
	>spacer3	
	TCTCTGTATAGAGTGTATTTATCGCTGTCCG	*Streptococcus* phages: 9A / vB_SthS_VA460 / P8922 / P7154 / P5652 / P5651 / P4761 / P0094 / P0093 / CHPC1151 / TP-778L; *Streptococcus thermophilus* bacteriophage 7201
	>spacer4	
	TAAGTTCACTACGTGCCATTTCAAGGCTAAG	
	>spacer5	
	ATAACGGTCAAAATATAGCTTACTTAGCTT	
	>spacer6	
	ATGCAATAAAATCGTGCTACGGGCGTTTTA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer7	
	TTTTTATTTTATAGTTTGCGTGGTCTAGTG	*Streptococcus* phages: MM25 / L5A1 / C0 / A0 / P9854 / P9853 / P9852 / P9851 / P8922 / P8921 / P7633 / P7602 / P7601 / P7574 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P3684 / P0094 / P0093 / P0092 / P0091 / 53 / CHPC1151; *Streptococcus thermophilus* bacteriophages: 2972 / 7201
	>spacer8	
	ATAACGGTCAAAATATAGCTTACTTAGCTT	
	>spacer9	
	ATGCAATAAAATCGTGCTACGGGCGTTTTA	*Streptococcus thermophilus* bacteriophage 7201

	>spacer10	
	GTATCGACGTTCGGACATATCCTAGCAAAA	*Streptococcus* phages: P7602 / P5641
	>spacer11	
	AGCGTAGCGTAGTGTATGTGTAGCGCCACC	*Streptococcus* phage P3681; *Streptococcus thermophilus* bacteriophage 7201
	>spacer12	
	CCACTAGCATCATATTCTTTTGACTTAGAA	*Streptococcus* phage P7633
	>spacer13	
	ATGTACTTCGAATTATTCAATTTTAAGAAG	*Streptococcus thermophilus* bacteriophage O1205
	>spacer14	
	TTAGGTCCAGGTACACCTTGGTCACCTCTAG	*Streptococcus* phages: P9852 / TP-778L / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer15	
	GGTCTTTCCGAATCCATTTATTAGATTGAG	*Streptococcus* phages: P98523 / P9852 / P7572 / P7571 / ALQ13.2 / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer16	
	AGGTGTGGCTACGGATTGAAACTGCTAACG	*Streptococcus* phages: V2 / STP2 / STP1 / R1 / MM25 / L5A1 / C0 / B5 / B0 / 31B4 / 9A / 7A5 / A0 / 7T / vB_SthS_VA698 / P9903 / P9902 / P9901 / P8922 / P7633 / P7632 / P7631 / P7602 / P7601 / P7574 / P7572 / P7571 / P7154 / P7133 / P5641 / P3684 / P3681 / P0095 / D4276 / 128 / 53 / CHPC926 / CHPC577
	>spacer17	
	CACAAAAGCCCAAAAATAGACAAGCCATTC	
	>spacer18	
	TTTAAGTTGTTTTATTAAGCATAGGCCGTT	*Streptococcus thermophilus* bacteriophage O1205
	>spacer19	
	TTTCTATGGTACACACCTAAACGAACACAA	*Streptococcus* phages: vB_SthS_VA214 / P5641 / P3684 / P3681 / P0091; *Streptococcus thermophilus* bacteriophage 7201
	>spacer20	

	GATAATCAAGGGGGCTTTTTTGATATTACT	*Streptococcus* phages: 9A / P9903 / P8922 / P7633 / P7632 / P7631 / P7154 / P7152 / P7151 / P5652 / P5651 / P3684
APC151 (575011.. 576303)	>spacer1	
	ATAAGATTGAAACATACTTGTGCAAGTGTG	*Streptococcus* phage D4276
	>spacer2	
	AATTATCTTGGTAGAAAACTTTATTATTTA	
	>spacer3	
	TCTCTGTATAGAGTGTATTTATCGCTGTCCG	*Streptococcus* phages: 9A / vB_SthS_VA460 / P8922 / P7154 / P5652 / P5651 / P4761 / P0094 / P0093 / CHPC1151 / TP-778L; *Streptococcus thermophilus* bacteriophage 7201
	>spacer4	
	TAAGTTCACTACGTGCCATTTCAAGGCTAAG	
	>spacer5	
	ATAACGGTCAAAATATAGCTTACTTAGCTT	
	>spacer6	
	TTTTTATTTTATAGTTTGCGTGGTCTAGTG	*Streptococcus* phages: MM25 / L5A1 / C0 / A0 / P9854 / P9853 / P9852 / P9851 / P8922 / P8921 / P7633 / P7602 / P7601 / P7574 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P3684 / P0094 / P0093 / P0092 / P0091 / 53 / CHPC1151; *Streptococcus thermophilus* bacteriophages: 2972 / 7201
	>spacer7	
	ATAACGGTCAAAATATAGCTTACTTAGCTT	
	>spacer8	
	ATGCAATAAAATCGTGCTACGGGCGTTTTA	*Streptococcus thermophilus* bacteriophage 7201
	>spacer9	
	GTATCGACGTTCGGACATATCCTAGCAAAA	*Streptococcus* phages: P7602 / P5641
	>spacer10	
	AGCGTAGCGTAGTGTATGTGTAGCGCCACC	*Streptococcus* phage P3681; *Streptococcus thermophilus* bacteriophage 7201
	>spacer11	
	CCACTAGCATCATATTCTTTTGACTTAGAA	*Streptococcus* phage P7633
	>spacer12	

	ATGTACTTCGAATTATTCAATTTTAAGAAG	*Streptococcus thermophilus* bacteriophage O1205
	>spacer13	
	TTAGGTCCAGGTACACCTTGGTCACCTCTAG	*Streptococcus* phages: P9852 / TP-778L / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer14	
	GGTCTTTCCGAATCCATTTATTAGATTGAG	*Streptococcus* phages: P98523 / P9852 / P7572 / P7571 / ALQ13.2 / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer15	
	AGGTGTGGCTACGGATTGAAACTGCTAACG	*Streptococcus* phages: V2 / STP2 / STP1 / R1 / MM25 / L5A1 / C0 / B5 / B0 / 31B4 / 9A / 7A5 / A0 / 7T / vB_SthS_VA698 / P9903 / P9902 / P9901 / P8922 / P7633 / P7632 / P7631 / P7602 / P7601 / P7574 / P7572 / P7571 / P7154 / P7133 / P5641 / P3684 / P3681 / P0095 / D4276 / 128 / 53 / CHPC926 / CHPC577
	>spacer16	
	CACAAAAGCCCAAAAATAGACAAGCCATTC	
	>spacer17	
	TTTAAGTTGTTTTATTAAGCATAGGCCGTT	*Streptococcus thermophilus* bacteriophage O1205
	>spacer18	
	TTTCTATGGTACACACCTAAACGAACACAA	*Streptococcus* phages: vB_SthS_VA214 / P5641 / P3684 / P3681 / P0091; *Streptococcus thermophilus* bacteriophage 7201
	>spacer19	
	GATAATCAAGGGGGCTTTTTTGATATTACT	*Streptococcus* phages: 9A / P9903 / P8922 / P7633 / P7632 / P7631 / P7154 / P7152 / P7151 / P5652 / P5651 / P3684
LMD-9 (1377229..1377794)	>spacer1	
	ACAATTATTTTTCAATGAAATTAATACTAT	
	>spacer2	
	TTTAAAGTAGTTTTATCTTGATAAAGTTCT	
	>spacer3	

	TTTATGACAAAAGCAACTAACGAATATGAG	*Streptococcus* phages: V2 / STP2 / STP1 / R1 / B5 / B0 / 31B4 / 16B8 / 9B4 /7A5 / 7T / vB_SthS_VA460 / vB_SthS_VA698 / P8922 / P7632 / P7631 / P7602 / P7154 / P7152 / P7151 / P5651 / P3684 / P3681 / 128; *Streptococcus thermophilus* bacteriophage Sfi19
	>spacer4	
	GTGTCTCCAATGAGGACAATATTGTACTCTGT	*Streptococcus* phage TP-J34
	>spacer5	
	TTAATCATGGCTACTGCGGTTTCCCAAACA	*Streptococcus* phages: M19 / A9 / P8922 / P7633 / P7632 / P7631 / P7154 / P7152 / P5652 / P5651 / P3684 / P3681 / 128; *Streptococcus thermophilus* bacteriophage 7201
	>spacer6	
	TATATGGGCGTATGGAGAGCCACGACAAGG	*Streptococcus* phages: 16B8 / 7T
	>spacer7	
	TAAGTTTTCGTCTTGGATACCACTCATTGA	
	>spacer8	
	TAAGTACTCGTACAGTGAACCTTTTTCACC	
SMQ-301 (1383838..1384867)	>spacer1	
	ACAATTATTTTTCAATGAAATTAATACTAT	
	>spacer2	
	TTTAAAGTAGTTTTATCTTGATAAAGTTCT	
	>spacer3	
	TTGCCTTGAGCTTGTATCACTTTTATTATT	*Streptococcus thermophilus* bacteriophage Sfi19
	>spacer4	
	GTAACAACTGAGTTAAAGTTAGAAGATTCA	*Streptococcus* phage 858
	>spacer5	
	TTTATGACAAAAGCAACTAACGAATATGAG	*Streptococcus* phages: V2 / STP2 / STP1 / R1 / B5 / B0 / 31B4 / 16B8 / 9B4 /7A5 / 7T / vB_SthS_VA460 / vB_SthS_VA698 / P8922 / P7632 / P7631 / P7602 / P7154 / P7152 / P7151 / P5651 / P3684 / P3681 / 128; *Streptococcus thermophilus* bacteriophage Sfi19
	>spacer6	

	GATGGCTTAAAAGGCACGCTCGTTTCAATA	
	>spacer7	
	CTATAGTTGTAATCTTCGCCATCGTAAGTAT	*Streptococcus* phages: vB_SthS_VA460 / P0091
	>spacer8	
	TACTCATTAGTAAAAACGCCTTTTTCCAAT	
	>spacer9	
	CTATAGTTGTAATCTTCGCCATCGTAAGTAT	*Streptococcus* phages: vB_SthS_VA460 / P0091
	>spacer10	
	TACTCATTAGTAAAAACGCCTTTTTCCAAT	
	>spacer11	
	GGTTTTCCTGGAATACCGTCTCGTCCATCA	*Streptococcus* phages: P9854 / P9853 / P9852
	>spacer12	
	GTGTCTCCAATGAGGACAATATTGTACTCTGT	*Streptococcus* phage TP-J34
	>spacer13	
	TTAATCATGGCTACTGCGGTTTCCCAAACA	*Streptococcus* phages: M19 / A9 / P8922 / P7633 / P7632 / P7631 / P7154 / P7152 / P5652 / P5651 / P3684 / P3681 / 128; *Streptococcus thermophilus* bacteriophage 7201
	>spacer14	
	TTACCAGTAAAATCATATCTAACCGTCAGA	*Streptococcus* phages: vB_SthS_VA214 / P0091
	>spacer15	
	TCATCAAAACATGAATATGAGGGTGATATG	*Streptococcus pneumoniae* plasmids: pPR3 / pSMB1 / pDP1
MN-ZLW-002 (1372720..1374472)	>spacer1	
	GCTCCACCAGAGGTAGACATATCAGCACCC	*Streptococcus* phage vB_SthS_VA460; *Streptococcus thermophilus* bacteriophage 7201
	>spacer2	
	GCCTGATAGTACAAGTCGTACAGCTTGTTA	*Streptococcus* phage D4276
	>spacer3	
	ATGTACTTCGTTAATATTCAATTTTTAAGA	*Streptococcus* phage vB_SthS_VA460
	>spacer4	
	TTACGTGACCCATACATTGAATAGGTCAAT	*Streptococcus* phages: CHPC926 / TP-778L / 858

	>spacer5	
	GTTCTTATGCGTGTTATAAGCTGAAAGTTC	
	>spacer6	
	CTATTCTGCGTTAGATTTCAAAGTTACCAG	
	>spacer7	
	AGCCAAGTTGTCTTTTTCGATACAGACCGT	
	>spacer8	
	ACATATCTAATGACATTCATGTCTGGGTTG	*Streptococcus* phage P0091
	>spacer9	
	TCGGTTTTCGACTTCATAAGTTTCTTGTTC	
	>spacer10	
	CTTCAATTATTTATTTGTCGTCCAGTGCTT	
	>spacer11	
	CTTGTTACACACATGTATTTAGATGATTTT	*Streptococcus* phages: MM25 / L5A1 / C0 / 31B4 / 7A5 / A0 / vB_SthS_VA698 / P9903 / P9902 / P9901 / P9853 / P9852 / P9851 / P7572 / P7571 / P7134 / P7132 / 53 / 73 / Abc2 / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer12	
	CTGACCGTTAGGATTTCCATATGCGTCAGG	
	>spacer13	
	GAGACCTCAGCGTCACCAGAGACCCTAGCG	*Streptococcus thermophilus* bacteriophages: Sfi18 / Sfi19
	>spacer14	
	TCCTTACCTGGCAAGCCATCATTTCCATCG	*Streptococcus* phage P0091
	>spacer15	
	CTGACTGTGAACGGTGTTCAGAAAAACACT	*Streptococcus* phages: P3684 / D4276
	>spacer16	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer17	
	ATACGACAACCATTCAAGAGCATAAACTCG	*Streptococcus* phage 5093
	>spacer18	
	GAGAAACTGAAGAACATTCCTGCTTTAAAT	*Streptococcus* phage P7602
	>spacer19	

	TTCAATGTGGTATTATTAGGGTTTTCACCC	*Streptococcus* phages: V2 / STP2 / STP1 / R1 / MM25 / M19 / L5A1 / C0 / B5 / B0 / 31B4 / 16B8 / 9B4 / 7A5 / A0 / 7T / vB_SthS_VA698 / P9903 / P9902 / P8922 / P8921 / P7602 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P5641 / P3681 / 128 / 53 / 73 / Abc2; *Streptococcus thermophilus* bacteriophage DT1
	>spacer20	
	TTTACAAGTGTCCATGAGTAATCGCTAGGA	*Streptococcus* phages: P9854 / P9852 / P9851 / P5641 / TP-778L / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer21	
	TTATTTCCTTCAGCTTTACCTTCAGAGATA	
	>spacer22	
	TCAGTTGAAAGGCCTTCTGGACCTTTAATA	*Streptococcus* phages: V2 / R1 / M19 / C0 / 31B4 / 16B8 / 9B4 / 9A / A0 / 7T / vB_SthS_VA460 / vB_SthS_VA214 / P9851 / P7633 / P7601 / P7132 / P5641 / P3684 / P3681 / 128 / 53 / 73
	>spacer23	
	TTCCGCACTCTTGTAATTGTCGAACGGTTT	
	>spacer24	
	TCACGACTAGTATATTCAGGTACACCGAAT	
	>spacer25	
	GTAGCAAGACTAAACACAACTGACGTATCT	
	>spacer26	
	CTCTGAGGACTGAGTAATGGACCAAGTGTTT	
MN-BM-A01 (1186347..1188099)	>spacer1	
	GCTCCACCAGAGGTAGACATATCAGCACCC	*Streptococcus* phage vB_SthS_VA460; *Streptococcus thermophilus* bacteriophage 7201
	>spacer2	
	GCCTGATAGTACAAGTCGTACAGCTTGTTA	*Streptococcus* phage D4276
	>spacer3	
	ATGTACTTCGTTAATATTCAATTTTTAAGA	*Streptococcus* phage vB_SthS_VA460
	>spacer4	
	TTACGTGACCCATACATTGAATAGGTCAAT	*Streptococcus* phages: CHPC926 / TP-778L / 858

	>spacer5	
	GTTCTTATGCGTGTTATAAGCTGAAAGTTC	
	>spacer6	
	CTATTCTGCGTTAGATTTCAAAGTTACCAG	
	>spacer7	
	AGCCAAGTTGTCTTTTTCGATACAGACCGT	
	>spacer8	
	ACATATCTAATGACATTCATGTCTGGGTTG	*Streptococcus* phage P0091
	>spacer9	
	TCGGTTTTCGACTTCATAAGTTTCTTGTTC	
	>spacer10	
	CTTCAATTATTTATTTGTCGTCCAGTGCTT	
	>spacer11	
	CTTGTTACACACATGTATTTAGATGATTTT	*Streptococcus* phages: MM25 / L5A1 / C0 / 31B4 / 7A5 / A0 / vB_SthS_VA698 / P9903 / P9902 / P9901 / P9853 / P9852 / P9851 / P7572 / P7571 / P7134 / P7132 / 53 / 73 / Abc2 / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer12	
	CTGACCGTTAGGATTTCCATATGCGTCAGG	
	>spacer13	
	GAGACCTCAGCGTCACCAGAGACCCTAGCG	*Streptococcus thermophilus* bacteriophages: Sfi18 / Sfi19
	>spacer14	
	TCCTTACCTGGCAAGCCATCATTTCCATCG	*Streptococcus* phage P0091
	>spacer15	
	CTGACTGTGAACGGTGTTCAGAAAAACACT	*Streptococcus* phages: P3684 / D4276
	>spacer16	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer17	
	ATACGACAACCATTCAAGAGCATAAACTCG	*Streptococcus* phage 5093
	>spacer18	
	GAGAAACTGAAGAACATTCCTGCTTTAAAT	*Streptococcus* phage P7602
	>spacer19	

	TTCAATGTGGTATTATTAGGGTTTTCACCC	*Streptococcus* phages: V2 / STP2 / STP1 / R1 / MM25 / M19 / L5A1 / C0 / B5 / B0 / 31B4 / 16B8 / 9B4 / 7A5 / A0 / 7T / vB_SthS_VA698 / P9903 / P9902 / P8922 / P8921 / P7602 / P7154 / P7152 / P7151 / P7134 / P7133 / P7132 / P5652 / P5651 / P5641 / P3681 / 128 / 53 / 73 / Abc2; *Streptococcus thermophilus* bacteriophage DT1
	>spacer20	
	TTTACAAGTGTCCATGAGTAATCGCTAGGA	*Streptococcus* phages: P9854 / P9852 / P9851 / P5641 / TP-778L / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer21	
	TTATTTCCTTCAGCTTTACCTTCAGAGATA	
	>spacer22	
	TCAGTTGAAAGGCCTTCTGGACCTTTAATA	*Streptococcus* phages: V2 / R1 / M19 / C0 / 31B4 / 16B8 / 9B4 / 9A / A0 / 7T / vB_SthS_VA460 / vB_SthS_VA214 / P9851 / P7633 / P7601 / P7132 / P5641 / P3684 / P3681 / 128 / 53 / 73
	>spacer23	
	TTCCGCACTCTTGTAATTGTCGAACGGTTT	
	>spacer24	
	TCACGACTAGTATATTCAGGTACACCGAAT	
	>spacer25	
	GTAGCAAGACTAAACACAACTGACGTATCT	
	>spacer26	
	CTCTGAGGACTGAGTAATGGACCAAGTGTTT	
ASCC 1275 (1558841..1559668)	>spacer1	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer2	
	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871
	>spacer3	
	TCTGCAATCTCTTCGGAAGTATAGTAATAG	*Streptococcus* phages: TP-778L / TP-J34 / *Streptococcus thermophilus* bacteriophages: Sfi21 / O1205
	>spacer4	

	AATACCGATTACACTATATCTGTTAAAATT	
	>spacer5	
	GCGCTACGCATTTGACTTTCTTGTGCCTTC	*Streptococcus* phages: P7574 / P7573
	>spacer6	
	CTCTGGATCATATATAGCTACTTTCCCATT	*Streptococcus* phages: MM25 / 9A / vB_SthS_VA460 / vB_SthS_VA698 / P9854 / P9853 / P9851 / P8922 / P8921 / P7954 / P7953 / P7952 / P7633 / P7632 / P7631 / P7601 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7133 / P7132 / P5652 / P5651 / P3684 / P3681 / 128 / ALQ13.2; *Streptococcus thermophilus* bacteriophages: kappa3 / 2972 / Sfi11 / DT1 / SFi18 / Sfi21 / Sfi19 / O1205
	>spacer7	
	TCCTTACCTGGCAAGCCATCATTTCCATCG	*Streptococcus* phage P0091
	>spacer8	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer9	
	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871
	>spacer10	
	ACATTATTAATAAAATGTTTACAAGAGGAA	
	>spacer11	
	CAAACCAACAACCAAGCTAATACAGCAGTA	*Streptococcus* phages: vB_SthS_VA460 / 9874
	>spacer12	
	TGTCCTCTTCCTCTTTAGCGTTTAGAATTT	*Streptococcus thermophilus* bacteriophage 7201
MN-BM-A02 (1368586..1369413)	>spacer1	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer2	
	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871
	>spacer3	
	TCTGCAATCTCTTCGGAAGTATAGTAATAG	*Streptococcus* phages: TP-778L / TP-J34 / *Streptococcus thermophilus* bacteriophages: Sfi21 / O1205

	>spacer4	
	AATACCGATTACACTATATCTGTTAAAATT	
	>spacer5	
	GCGCTACGCATTTGACTTTCTTGTGCCTTC	*Streptococcus* phages: P7574 / P7573
	>spacer6	
	CTCTGGATCATATATAGCTACTTTCCCATT	*Streptococcus* phages: MM25 / 9A / vB_SthS_VA460 / vB_SthS_VA698 / P9854 / P9853 / P9851 / P8922 / P8921 / P7954 / P7953 / P7952 / P7633 / P7632 / P7631 / P7601 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7133 / P7132 / P5652 / P5651 / P3684 / P3681 / 128 / ALQ13.2; *Streptococcus thermophilus* bacteriophages: kappa3 / 2972 / Sfi11 / DT1 / SFi18 / Sfi21 / Sfi19 / O1205
	>spacer7	
	TCCTTACCTGGCAAGCCATCATTTCCATCG	*Streptococcus* phage P0091
	>spacer8	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer9	
	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871
	>spacer10	
	ACATTATTAATAAAATGTTTACAAGAGGAA	
	>spacer11	
	CAAACCAACAACCAAGCTAATACAGCAGTA	*Streptococcus* phages: vB_SthS_VA460 / 9874
	>spacer12	
	TGTCCTCTTCCTCTTTAGCGTTTAGAATTT	*Streptococcus thermophilus* bacteriophage 7201
KLDS SM (683810..684637)	>spacer1	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer2	
	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871
	>spacer3	

	TCTGCAATCTCTTCGGAAGTATAGTAATAG	*Streptococcus* phages: TP-778L / TP-J34 / *Streptococcus thermophilus* bacteriophages: Sfi21 / O1205
	>spacer4	
	AATACCGATTACACTATATCTGTTAAAATT	
	>spacer5	
	GCGCTACGCATTTGACTTTCTTGTGCCTTC	*Streptococcus* phages: P7574 / P7573
	>spacer6	
	CTCTGGATCATATATAGCTACTTTCCCATT	*Streptococcus* phages: MM25 / 9A / vB_SthS_VA460 / vB_SthS_VA698 / P9854 / P9853 / P9851 / P8922 / P8921 / P7954 / P7953 / P7952 / P7633 / P7632 / P7631 / P7601 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7133 / P7132 / P5652 / P5651 / P3684 / P3681 / 128 / ALQ13.2; *Streptococcus thermophilus* bacteriophages: kappa3 / 2972 / Sfi11 / DT1 / SFi18 / Sfi21 / Sfi19 / O1205
	>spacer7	
	TCCTTACCTGGCAAGCCATCATTTCCATCG	*Streptococcus* phage P0091
	>spacer8	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer9	
	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871
	>spacer10	
	ACATTATTAATAAAATGTTTACAAGAGGAA	
	>spacer11	
	CAAACCAACAACCAAGCTAATACAGCAGTA	*Streptococcus* phages: vB_SthS_VA460 / 9874
	>spacer12	
	TGTCCTCTTCCTCTTTAGCGTTTAGAATTT	*Streptococcus thermophilus* bacteriophage 7201
ND07 (1294523..1295350) (rev_com)	>spacer1	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer2	
	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871

	>spacer3	
	TCTGCAATCTCTTCGGAAGTATAGTAATAG	*Streptococcus* phages: TP-778L / TP-J34 / *Streptococcus thermophilus* bacteriophages: Sfi21 / O1205
	>spacer4	
	AATACCGATTACACTATATCTGTTAAAATT	
	>spacer5	
	GCGCTACGCATTTGACTTTCTTGTGCCTTC	*Streptococcus* phages: P7574 / P7573
	>spacer6	
	CTCTGGATCATATATAGCTACTTTCCCATT	*Streptococcus* phages: MM25 / 9A / vB_SthS_VA460 / vB_SthS_VA698 / P9854 / P9853 / P9851 / P8922 / P8921 / P7954 / P7953 / P7952 / P7633 / P7632 / P7631 / P7601 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7133 / P7132 / P5652 / P5651 / P3684 / P3681 / 128 / ALQ13.2; *Streptococcus thermophilus* bacteriophages: kappa3 / 2972 / Sfi11 / DT1 / SFi18 / Sfi21 / Sfi19 / O1205
	>spacer7	
	TCCTTACCTGGCAAGCCATCATTTCCATCG	*Streptococcus* phage P0091
	>spacer8	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer9	
	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871
	>spacer10	
	ACATTATTAATAAAATGTTTACAAGAGGAA	
	>spacer11	
	CAAACCAACAACCAAGCTAATACAGCAGTA	*Streptococcus* phages: vB_SthS_VA460 / 9874
	>spacer12	
	TGTCCTCTTCCTCTTTAGCGTTTAGAATTT	*Streptococcus thermophilus* bacteriophage 7201
DGCC 7710 (1371965..1372792)	>spacer1	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer2	

	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871
	>spacer3	
	TCTGCAATCTCTTCGGAAGTATAGTAATAG	*Streptococcus* phages: TP-778L / TP-J34 / *Streptococcus thermophilus* bacteriophages: Sfi21 / O1205
	>spacer4	
	AATACCGATTACACTATATCTGTTAAAATT	
	>spacer5	
	GCGCTACGCATTTGACTTTCTTGTGCCTTC	*Streptococcus* phages: P7574 / P7573
	>spacer6	
	CTCTGGATCATATATAGCTACTTTCCCATT	*Streptococcus* phages: MM25 / 9A / vB_SthS_VA460 / vB_SthS_VA698 / P9854 / P9853 / P9851 / P8922 / P8921 / P7954 / P7953 / P7952 / P7633 / P7632 / P7631 / P7601 / P7573 / P7572 / P7571 / P7154 / P7152 / P7151 / P7133 / P7132 / P5652 / P5651 / P3684 / P3681 / 128 / ALQ13.2; *Streptococcus thermophilus* bacteriophages: kappa3 / 2972 / Sfi11 / DT1 / SFi18 / Sfi21 / Sfi19 / O1205
	>spacer7	
	TCCTTACCTGGCAAGCCATCATTTCCATCG	*Streptococcus* phage P0091
	>spacer8	
	TTAGCGCAACCGTTAATTTGACAATTCATA	*Streptococcus* phages: CHPC926 / CHPC577 / 9874 / 9873 / 9872 / 9871
	>spacer9	
	TCTATAATTGTTCTTCCATTCTCTGGGATA	*Streptococcus* phages: P7152 / P7134 / P0094 / P0093 / P0091 / 9871
	>spacer10	
	ACATTATTAATAAAATGTTTACAAGAGGAA	
	>spacer11	
	CAAACCAACAACCAAGCTAATACAGCAGTA	*Streptococcus* phages: vB_SthS_VA460 / 9874
	>spacer12	
	TGTCCTCTTCCTCTTTAGCGTTTAGAATTT	*Streptococcus thermophilus* bacteriophage 7201
GABA (1396122..1397147)	>spacer1	
	ACGCCCCCGAAAACGTAACAGGTAGCCCGC	

	>spacer2	
	ACGGTGCAAGCTTTCTTGGGTACGAATACC	*Streptococcus* phage D4276
	>spacer3	
	ATCTAGTCTCATAAAGTAAAACACTTTCTA	
	>spacer4	
	TGTCCAATTTGTACTCATGTAATAACACAA	
	>spacer5	
	GCAATTGAAGACATACAAGATGAAATTGAT	*Streptococcus* phages: P7955 / P7953 / P7952 / P7951
	>spacer6	
	GCTTTGGTTAAAGGGAAATACCTTTGGACT	*Streptococcus* phage P4761
	>spacer7	
	ATCCTTCTCCCTCTTGCTTGAGAGTAATCA	
	>spacer8	
	CTAAAAGAAACATAGATATTATCAAATTCA	
	>spacer9	
	CTTTACCCTTAAGCAACATAAAAAACAATG	
	>spacer10	
	CTGTTTGTGGTATCTAGCCTAAACAGGTTT	*Streptococcus* phage ALQ13.2
	>spacer11	
	TCACCAGTGTCTGTAATGTTGTTTTCATCA	Streptococcus phages: STP2 / STP1
	>spacer12	
	AGGCCAAAACCCCACAAATACATAAAAGAG	
	>spacer13	
	CAATATAAAGCCCCTCATAAAGTCCACCGT	
	>spacer14	
	TTCTCGACTTTAGTAAGATAATTCCTAACA	
	>spacer15	
	TACCTGAACTAGATAGCAAAATGGACGTCT	
ST3 (1399223..1400777)	>spacer1	
	TCTCTGTATAGAGTGTATTTATCGCTGTCCG	*Streptococcus* phages: 9A / vB_SthS_VA460 / P8922 / P7154 / P5652 / P5651 / P4761 / P0094 / P0093 / CHPC1151 / TP-778L; *Streptococcus thermophilus* bacteriophage 7201
	>spacer2	

	TGTCGCTCATGTGGTTAGCTGTGTCAGATT	*Streptococcus* phages: P9853 / P9852 / P7572 / P7571 / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer3	
	TCTTGGATATTGTGTGGAATGTTAAGCCCG	*Streptococcus* phages: P9854 / P9851 / 9874
	>spacer4	
	TAGTAGCCGCCATCGCCCATGTCAGTGATG	*Streptococcus thermophilus* bacteriophage 7201
	>spacer5	
	TTACTTTCAAGATAATCCTCAACCATTGAG	
	>spacer6	
	TGAGGGCCTTGTGGTCCTTGAATCCCTTGG	
	>spacer7	
	AACTATTTAGGGCAAGTTGCCAGGGAGGTA	*Streptococcus* phage D4276
	>spacer8	
	ACACGCTTGACGAAGAAATAAACGCATGGA	*Streptococcus* phages: P7602 / P7601
	>spacer9	
	ACGGTGGCTTTTGAATTTAACAACCTTGAT	*Streptococcus thermophilus* bacteriophage 7201
	>spacer10	
	TTATAACTTTTACATTTATTACACCTAACA	*Streptococcus* phages: vB_SthS_VA460 / P7632 / P7631
	>spacer11	
	ACAGGTTTGTAGGTTATTTCCATAAGCGTC	
	>spacer12	
	CGGTAATACGCATCCACTGAATCTTAGGGT	*Streptococcus* phages: M19 / vB_SthS_VA460 / vB_SthS_VA460 / P9902 / P9853 / CHPC577 / 9874 / 9873 / 9872 / 9871 / 858; *Streptococcus thermophilus* bacteriophage 2972
	>spacer13	
	ATTTTTTCGGCTAACTCGACTCTTGTCATT	*Streptococcus* phage M19
	>spacer14	
	TGCTTTAAGACTGGTTTAACAGTAGAACGA	*Streptococcus thermophilus* bacteriophages: DT1.1 / DT1.4 / MD4 / 7201
	>spacer15	
	CGACCATATGAAGACAATGTAACTGTATTA	*Streptococcus* phages: P7571 / P7132
	>spacer16	

	AAACCAAACTGGCAAATTGCCCAAGCAATC	
	>spacer17	
	CTGAGCTTTCACTAATCGAACCAATCCCAT	*Streptococcus* phages: P7955 / P7954 / P7953 / P7952 / TP-778L / TP-J34 / ALQ13.2; *Streptococcus thermophilus* bacteriophage Sfi11
	>spacer18	
	AAATAATCGGCTAACTTCTTAGCATTACTT	*Streptococcus* phages: P0092 / P0091 / CHPC1151 / TP-J34
	>spacer19	
	CTTACATCATGGATGTTAAAGGGTGCAAAT	*Streptococcus* phages: 9A / vB_SthS_VA460 / P9902 / P9854 / P9852 / P9851 / P7152 / P7151 / P5652 / P0095 / 53 / 9874 / Abc2; *Streptococcus thermophilus* bacteriophages: DT1 / 7201
	>spacer20	
	TTAACCATATCAGTATGCATTGAACCAGAT	*Streptococcus* phages: vB_SthS_VA698 / D4276 / ALQ13.2; *Streptococcus thermophilus* bacteriophage 7201
	>spacer21	
	CGACTTTTGTAGATATGGAAGTTGCGTATA	*Streptococcus* phages: vB_SthS_VA698 / P7574 / P5641
	>spacer22	
	GTAGAACTGTACGACATCATCAAACCTTAG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer23	
	CTACCACCCCGTGGCATAAATGTGAATTCC	*Streptococcus* phages: P0094 / P0093 / P0092 / CHPC1151 / 5093
NCTC12958 (1568082..1569571)	>spacer1	
	ACCATTCGTTGGTGCTACCGTTCCAGATGT	*Streptococcus* phage D4276; *Streptococcus thermophilus* bacteriophage 7201
	>spacer2	
	GCTTATGCTAGAAGAACTAGCGAAGAACGT	
	>spacer3	
	GTTTGGGTATGAGGTTGAGGAAGAAATGAA	
	>spacer4	
	ATTTCTTTAATATTCCATCTACTACCTGTT	
	>spacer5	
	ATAAGATTGAAACATACTTGTGCAAGTGTG	*Streptococcus* phage D4276

	>spacer6	
	TTCCAACCGATGAAGTTATTATGACCTATG	*Streptococcus* phages: P8922 / P7602 / P7601 / P7154 / P7152; *Streptococcus thermophilus* bacteriophags: Sfi21 / Sfi19
	>spacer7	
	TACTTAAGACCATTAAGAGGGTAACTCTTAAG	
	>spacer8	
	TAATAACCACCGTTTCCGGTATCTGTGATG	
	>spacer9	
	AATTATTTTGTGGCATATCAAAACTAATAA	*Streptococcus* phages: D1811 / P8921 / P7633; *Streptococcus thermophilus* bacteriophage DT2
	>spacer10	
	CTTATCTCGAGATACTCAATTCAATTGAAG	
	>spacer11	
	TTGGGGCAGCCTTGCAAGGCATTGCTAGTG	*Streptococcus* phages: P4761 / P0095 / P0094 / P0093 / P0092 / 5093
	>spacer12	
	AGTTAACCATATCAACAGGAATTCCATTTT	*Streptococcus* phages: M19 / 53
	>spacer13	
	GCCTGATAGTACAAGTCGTACAGCTTGTTA	*Streptococcus* phage D4276
	>spacer14	
	TCCGCTTCATCATTTCTGTTTTTGTTTGGC	*Streptococcus* phage D4276
	>spacer15	
	GGAAGCTACGGACACACAAGTGTGTTAGGT	*Streptococcus* phages: P7955 / P7951
	>spacer16	
	TTCCCTGTGTCTGTGATGTTGACTTCATCA	*Streptococcus* phages: D5842 / MM25 / vB_SthS_VA698 / P9854 / P9853 / P9852 / P9851 / P7632 / P7631 / P7134 / P7133 / P0094 / P0093 / P0092 / 128; *Streptococcus thermophilus* bacteriophage 2972
	>spacer17	
	ATTCACCCATACTGCCGCTGTGCAGTCGTT	*Streptococcus* phages: P7951 / P7572 / P7571 / P4761 / TP-778L; *Streptococcus thermophilus* bacteriophages: Sfi11 / O1205
	>spacer18	
	CTGACCGTTAGGATTTCCATATGCGTCAGG	
	>spacer19	

	TTATTTTCTCTGGACTGTTGAAGTCTTTCT	*Streptococcus thermophilus* bacteriophage Sfi21
	>spacer20	
	AAAAGAAAGCTCGTAAAATCGTTGAATTAA	
	>spacer21	
	ATGCTATTCGGATTCGGTATCTCAACTGTG	*Streptococcus* phages: vB_SthS_VA214 / P0095
	>spacer22	
	GCCACGTCAAGAGTGCCAAACTCACGTATT	*Streptococcus* phages: P7953 / P7955 / P7952
Spacers of CRISPR4		
Strain (coordinates of CRISPR array)	**spacers**	**1st blastn result (excluding same species)**
ASCC 1275 (1144778..1145540)	>spacer1	
	TGTTCCCATTGAGCCAATGCCAAACACACGACCA	*Streptococcus* phages: P7954 / P7953 / P7952
	>spacer2	
	TTTGAAAGGGGTTAAGGGTAAGGAAGTTGAGTAT	
	>spacer3	
	GTGGCACACCTACCTTGAATACTCAACCAGTACA	*Streptococcus* phages: 9A / P8922 / P7632 / P7631 / P7154 / P7152 / P7151 / P5652 / P5651 / 128; *Streptococcus thermophilus* bacteriophage 7201
	>spacer4	
	TTCTGTTCTTCTTGATGCTTGTCGAGGCGCTGA	*Streptococcus* phages: P4761 / TP-778L / TP-J34; *Streptococcus thermophilus* bacteriophage DT1
	>spacer5	
	CAATGTTCTCAGCTGGAAATGAGGTAATGTATA	*Streptococcus thermophilus* bacteriophages: MD2 / Sfi21 / Sfi19
	>spacer6	
	CAACGTTCCAACATTGTCGGCAATTTTACCTTG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer7	
	ATATTGTTAATGATAACTACAAATTTAAACGAG	*Streptococcus* phages: P7953 / V2 / STP2 / R1 / L5A1 / C0 / B5 / B0 / 31B4 / 9B4 / A0 / 7T / vB_SthS_VA214 / vB_SthS_VA698 / P8921 / P7951 / P7152 / P7132 / P5651 / P3681; *Streptococcus thermophilus* bacteriophages: Sfi11 / Sfi21 / Sfi19 / O1205
	>spacer8	
	CAAACGACAAAGGTCTATCAGCCTTGTAAGGGT	*Streptococcus* phage ALQ13.2

	>spacer9	
	CAGAAGAATCTGATCTATTCGCCGATAATACAG	*Streptococcus* phages: P7572 / P7571
	>spacer10	
	TAGTACTGTTCCCATGTCTAAGGAGGGGTTGCG	*Streptococcus* phage TP-778L
	>spacer11	
	CTGGGGGCGCTCAGACTCGCCCTAGTGGTGTCG	
	>spacer12	
	GCTTATGCGTGGGAGGCCATTGATATAGGTATA	*Streptococcus* phages: 9A / P8922 / P8921 / P7953 / P7633 / P7601 / P7154 / P7152 / P7151 / P5652 / P0095 / ALQ13.2
MN-BM-A02 (954527..955289)	>spacer1	
	TGTTCCCATTGAGCCAATGCCAAACACACGACCA	*Streptococcus* phages: P7954 / P7953 / P7952
	>spacer2	
	TTTGAAAGGGGTTAAGGGTAAGGAAGTTGAGTAT	
	>spacer3	
	GTGGCACACCTACCTTGAATACTCAACCAGTACA	*Streptococcus* phages: 9A / P8922 / P7632 / P7631 / P7154 / P7152 / P7151 / P5652 / P5651 / 128; *Streptococcus thermophilus* bacteriophage 7201
	>spacer4	
	TTCTGTTCTTCTTGATGCTTGTCGAGGCGCTGA	*Streptococcus* phages: P4761 / TP-778L / TP-J34; *Streptococcus thermophilus* bacteriophage DT1
	>spacer5	
	CAATGTTCTCAGCTGGAAATGAGGTAATGTATA	*Streptococcus thermophilus* bacteriophages: MD2 / Sfi21 / Sfi19
	>spacer6	
	CAACGTTCCAACATTGTCGGCAATTTTACCTTG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer7	
	ATATTGTTAATGATAACTACAAATTTAAACGAG	*Streptococcus* phages: P7953 / V2 / STP2 / R1 / L5A1 / C0 / B5 / B0 / 31B4 / 9B4 / A0 / 7T / vB_SthS_VA214 / vB_SthS_VA698 / P8921 / P7951 / P7152 / P7132 / P5651 / P3681; *Streptococcus thermophilus* bacteriophages: Sfi11 / Sfi21 / Sfi19 / O1205
	>spacer8	
	CAAACGACAAAGGTCTATCAGCCTTGTAAGGGT	*Streptococcus* phage ALQ13.2
	>spacer9	

	CAGAAGAATCTGATCTATTCGCCGATAATACAG	*Streptococcus* phages: P7572 / P7571
	>spacer10	
	TAGTACTGTTCCCATGTCTAAGGAGGGGTTGCG	*Streptococcus* phage TP-778L
	>spacer11	
	CTGGGGGCGCTCAGACTCGCCCTAGTGGTGTCG	
	>spacer12	
	GCTTATGCGTGGGAGGCCATTGATATAGGTATA	*Streptococcus* phages: 9A / P8922 / P8921 / P7953 / P7633 / P7601 / P7154 / P7152 / P7151 / P5652 / P0095 / ALQ13.2
KLDS SM (269742..270504)	>spacer1	
	TGTTCCCATTGAGCCAATGCCAAACACACGACCA	*Streptococcus* phages: P7954 / P7953 / P7952
	>spacer2	
	TTTGAAAGGGGTTAAGGGTAAGGAAGTTGAGTAT	
	>spacer3	
	GTGGCACACCTACCTTGAATACTCAACCAGTACA	*Streptococcus* phages: 9A / P8922 / P7632 / P7631 / P7154 / P7152 / P7151 / P5652 / P5651 / 128; *Streptococcus thermophilus* bacteriophage 7201
	>spacer4	
	TTCTGTTCTTCTTGATGCTTGTCGAGGCGCTGA	*Streptococcus* phages: P4761 / TP-778L / TP-J34; *Streptococcus thermophilus* bacteriophage DT1
	>spacer5	
	CAATGTTCTCAGCTGGAAATGAGGTAATGTATA	*Streptococcus thermophilus* bacteriophages: MD2 / Sfi21 / Sfi19
	>spacer6	
	CAACGTTCCAACATTGTCGGCAATTTTACCTTG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer7	
	ATATTGTTAATGATAACTACAAATTTAAACGAG	*Streptococcus* phages: P7953 / V2 / STP2 / R1 / L5A1 / C0 / B5 / B0 / 31B4 / 9B4 / A0 / 7T / vB_SthS_VA214 / vB_SthS_VA698 / P8921 / P7951 / P7152 / P7132 / P5651 / P3681; *Streptococcus thermophilus* bacteriophages: Sfi11 / Sfi21 / Sfi19 / O1205
	>spacer8	
	CAAACGACAAAGGTCTATCAGCCTTGTAAGGGT	*Streptococcus* phage ALQ13.2
	>spacer9	
	CAGAAGAATCTGATCTATTCGCCGATAATACAG	*Streptococcus* phages: P7572 / P7571

	>spacer10	
	TAGTACTGTTCCCATGTCTAAGGAGGGGTTGCG	*Streptococcus* phage TP-778L
	>spacer11	
	CTGGGGGCGCTCAGACTCGCCCTAGTGGTGTCG	
	>spacer12	
	GCTTATGCGTGGGAGGCCATTGATATAGGTATA	*Streptococcus* phages: 9A / P8922 / P8921 / P7953 / P7633 / P7601 / P7154 / P7152 / P7151 / P5652 / P0095 / ALQ13.2
ND07 (1708655..1709417) (rev_com)	>spacer1	
	TGTTCCCATTGAGCCAATGCCAAACACACGACCA	*Streptococcus* phages: P7954 / P7953 / P7952
	>spacer2	
	TTTGAAAGGGGTTAAGGGTAAGGAAGTTGAGTAT	
	>spacer3	
	GTGGCACACCTACCTTGAATACTCAACCAGTACA	*Streptococcus* phages: 9A / P8922 / P7632 / P7631 / P7154 / P7152 / P7151 / P5652 / P5651 / 128; *Streptococcus thermophilus* bacteriophage 7201
	>spacer4	
	TTCTGTTCTTCTTGATGCTTGTCGAGGCGCTGA	*Streptococcus* phages: P4761 / TP-778L / TP-J34; *Streptococcus thermophilus* bacteriophage DT1
	>spacer5	
	CAATGTTCTCAGCTGGAAATGAGGTAATGTATA	*Streptococcus thermophilus* bacteriophages: MD2 / Sfi21 / Sfi19
	>spacer6	
	CAACGTTCCAACATTGTCGGCAATTTTACCTTG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer7	
	ATATTGTTAATGATAACTACAAATTTAAACGAG	*Streptococcus* phages: P7953 / V2 / STP2 / R1 / L5A1 / C0 / B5 / B0 / 31B4 / 9B4 / A0 / 7T / vB_SthS_VA214 / vB_SthS_VA698 / P8921 / P7951 / P7152 / P7132 / P5651 / P3681; *Streptococcus thermophilus* bacteriophages: Sfi11 / Sfi21 / Sfi19 / O1205
	>spacer8	
	CAAACGACAAAGGTCTATCAGCCTTGTAAGGGT	*Streptococcus* phage ALQ13.2
	>spacer9	
	CAGAAGAATCTGATCTATTCGCCGATAATACAG	*Streptococcus* phages: P7572 / P7571
	>spacer10	

	TAGTACTGTTCCCATGTCTAAGGAGGGGTTGCG	*Streptococcus* phage TP-778L
	>spacer11	
	CTGGGGGCGCTCAGACTCGCCCTAGTGGTGTCG	
	>spacer12	
	GCTTATGCGTGGGAGGCCATTGATATAGGTATA	*Streptococcus* phages: 9A / P8922 / P8921 / P7953 / P7633 / P7601 / P7154 / P7152 / P7151 / P5652 / P0095 / ALQ13.2
DGCC 7710 (957897..958659)	>spacer1	
	TGTTCCCATTGAGCCAATGCCAAACACACGACCA	*Streptococcus* phages: P7954 / P7953 / P7952
	>spacer2	
	TTTGAAAGGGGTTAAGGGTAAGGAAGTTGAGTAT	
	>spacer3	
	GTGGCACACCTACCTTGAATACTCAACCAGTACA	*Streptococcus* phages: 9A / P8922 / P7632 / P7631 / P7154 / P7152 / P7151 / P5652 / P5651 / 128; *Streptococcus thermophilus* bacteriophage 7201
	>spacer4	
	TTCTGTTCTTCTTGATGCTTGTCGAGGCGCTGA	*Streptococcus* phages: P4761 / TP-778L / TP-J34; *Streptococcus thermophilus* bacteriophage DT1
	>spacer5	
	CAATGTTCTCAGCTGGAAATGAGGTAATGTATA	*Streptococcus thermophilus* bacteriophages: MD2 / Sfi21 / Sfi19
	>spacer6	
	CAACGTTCCAACATTGTCGGCAATTTTACCTTG	*Streptococcus thermophilus* bacteriophages: Sfi21 / Sfi19
	>spacer7	
	ATATTGTTAATGATAACTACAAATTTAAACGAG	*Streptococcus* phages: P7953 / V2 / STP2 / R1 / L5A1 / C0 / B5 / B0 / 31B4 / 9B4 / A0 / 7T / vB_SthS_VA214 / vB_SthS_VA698 / P8921 / P7951 / P7152 / P7132 / P5651 / P3681; *Streptococcus thermophilus* bacteriophages: Sfi11 / Sfi21 / Sfi19 / O1205
	>spacer8	
	CAAACGACAAAGGTCTATCAGCCTTGTAAGGGT	*Streptococcus* phage ALQ13.2
	>spacer9	
	CAGAAGAATCTGATCTATTCGCCGATAATACAG	*Streptococcus* phages: P7572 / P7571
	>spacer10	
	TAGTACTGTTCCCATGTCTAAGGAGGGGTTGCG	*Streptococcus* phage TP-778L

	>spacer11	
	CTGGGGGCGCTCAGACTCGCCCTAGTGGTGTCG	
	>spacer12	
	GCTTATGCGTGGGAGGCCATTGATATAGGTATA	*Streptococcus* phages: 9A / P8922 / P8921 / P7953 / P7633 / P7601 / P7154 / P7152 / P7151 / P5652 / P0095 / ALQ13.2
B59671 (295226..295498)	>spacer1	
	GCTTTTCGTTTAGCATCACTAACGCTTAATCCA	*Streptococcus* phages: ALQ13.2 / P9853 / P9852 / P7955 / P7954 / P7953 / P7952 / TP-J34 / 858; *Streptococcus thermophilus* bacteriophages: 2972 / O1205
	>spacer2	
	CAATCCTGATTACGACTTATACGATAGAAGCCA	*Streptococcus* phages: P7152 / P7132
	>spacer3	
	TCAAACTTTGATTAAACGTCAAGTTGGTGGTATG	*Streptococcus* phages: V2 / STP2 / STP1 / R1 / L5A1 / C0 / B5 / B0 / 31B4 / 16B8 / 9B4 / 9A / 7A5 / A0 / 7T / vB_SthS_VA698 / P9903 / P7632 / P7631 / P7602 / P7134 / P7133 / P5641 / 128 / Abc2; *Streptococcus thermophilus* bacteriophages: MD2 / DT1
	>spacer4	
	AAAAAAGTAATTTCTTCTGAAGCTACTTCTTTA	*Streptococcus* phages: P8921 / P7602 / D4276; *Streptococcus thermophilus* bacteriophages: DT2 / Sfi21 / Sfi19
Spacers of orphan CRISPR arrays		
Strain (coordinates of CRISPR array)	**spacers**	
JIM 8232 (1452282..1452514)	>spacer1	
	TCTCTCTCATAGATATCCTTGTGTTCTTTC	
	>spacer2	
	ATCGTCAATACAGCGACCAGCGACCACGA	
	>spacer3	
	TTCTCTTGCTACCAGCTTCAAGGCTTTTTT	
LMG 18311 (774319..774515)	>spacer1	
	GTCCTCCATGATTCAT	
	>spacer2	
	CATGCATATTGTCCATATCATG	
	>spacer3	
	TATCCATATCATGCATATCATG	

	>spacer4
	TGTGCATGTTGTGCATATTGTC

Supplementary Table S10. Restriction-modification (R-M) systems predicted in the 23 *S. thermophilus* strains after manual curation. Dashed lines are used to separate the two major clusters (A and B) and strain NCTC12958^T of the species, as described in the text

	Number of predicted R-M systems				
Strain	**Type I**	**Type II**	**Type III**	**Type IV**	**Total**
NCTC12958^T	2	2 (2)	1	0	5
KLDS 3.1003	2	1 (1)	0	1	4
ASCC 1275	4 (3)[1]	0	1 (1)	1	6
ND07	3 (2)	0	1 (1)	1	5
DGCC 7710	3 (2)	0	0	1	4
KLDS SM	3 (2)	0	1 (1)	1	5
MN-BM-A02	3 (2)	0	1 (1)	1	5
MN-ZLW-002	3 (3)	2	1 (1)	0	6
MN-BM-A01	2 (2)	3 (1)	0	0	5
JIM 8232	4 (3)	2 (2)	1 (1)	1	8
LMD-9	3 (2)	3 (1)	0	0	6
SMQ-301	2 (1)	2 (1)	0	0	4
ND03	2 (1)	3 (1)	0	0	5
APC151	1	3 (1)	0	0	4
GABA	2	3 (2)	1	0	6
ST3	1 (1)	1	0	1	3
CNRZ1066	3 (2)	3 (2)	1	1 (1)	8
CS8	3 (2)	2 (1)	1	0	6
S9	1	1 (1)	1	0	3
EPS	2	1 (1)	1	0	4
LMG 18311	4 (3)	2 (1)	1	1 (1)	8
B59671	0	2 (1)	1	0	3
ACA-DC 2	1	1 (1)	1	1	4

[1]Number in parentheses represent potentially inactivated R-M systems due to the presence of putative pseudogenes or the absence of one or more required protein subunits

Supplementary Table S11. General characteristics of integrated genomic islands (GIs) identified in the 23 S. thermophilus strains. Dashed lines are used to separate the two major clusters (A and B) and strain NCTC12958^T of the species, as described in the text

Strain	Predicted GIs	Unique GIs	Size range (bp)	GC (%) range	CDSs / pseudogenes	GIs excluded from analysis[1]
NCTC12958^T	23	11	3,555-38,494	30,3-42,9	281/30	GI 22
KLDS 3.1003	12	1	4,378-22,858	26,5-39,1	111/37	-
ASCC 1275	11	0	5,578-35,763	26,4-42,1	140/28	GI 1
ND07	12	0	4,070-38,272	26,1-42,3	144/28	GI 7
DGCC 7710	13	0	4,070-39,057	26,4-39,0	138/23	GI 12
KLDS SM	13	1	4,070-38,272	26,4-42.1	152/19	GI 7
MN-BM-A02	13	0	4,070-39,601	26,4-42,1	156/22	GI 12
MN-ZLW-002	10	0	4,411-48,230	30,1-42,9	144/25	-
MN-BM-A01	10	1	4,411-48,298	30,1-42,9	140/28	-
JIM 8232	12	6	4,477-58,211	26,5-41,6	164/7	GI 12
LMD-9	10	1	4,426-39,145	26,4-40,9	140/13	GI 10
SMQ-301	9	0	4,426-38,272	26,4-39,5	128/13	GI 9
ND03	10	0	5,091-22,592	29,3-39,7	141/24	GI 9
APC151	10	0	5,091-22,592	29,4-39,8	136/24	GI 6
GABA	12	4	3,555-39,209	26,4-45,2	137/16	GI 12
ST3	7	1	5,607-39,428	26,4-37,8	101/20	GI 7
CNRZ1066	11	0	4,087-25,706	29,9-39,7	118/31	-
CS8	11	0	4,411-21,101	26,5-39,4	107/26	-
S9	10	0	4,109-21,101	26,5-40,3	93/24	-
EPS	9	0	4,333-15,950	26,5-41,0	63/14	-
LMG 18311	10	1	4,591-36,887	30,1-40,5	139/17	GI 9
B59671	5	2	4,173-10,973	30,2-40,6	35/8	-
ACA-DC 2	10	2	4,150-14,221	30,1-38,9	72/18	-

The respective GIs include the array of ribosomal proteins and as false positive results were excluded from further analysis

Supplementary Table S12A. Genes in the predicted genomic islands (GIs) identified in the 23 *S. thermophilus* strains

Strain ACA-DC 2						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	14,221	100,859	115,080	33.1	STACADC2_RS00595	hypothetical protein
					STACADC2_RS00600	ketopantoate reductase family protein
					STACADC2_RS00605	50S ribosomal protein L13
					STACADC2_RS00610	30S ribosomal protein S9
					STACADC2_RS00615	hypothetical protein
					STACADC2_RS00620	pseudo
					STACADC2_RS00625	helix-turn-helix transcriptional regulator
					STACADC2_RS00630	hypothetical protein
					STACADC2_RS00635	pseudo
					STACADC2_RS00640	lantibiotic biosynthesis protein
					STACADC2_RS00645	lantibiotic transporter
					STACADC2_RS00650	pseudo
					STACADC2_RS00655	pseudo
					STACADC2_RS00660	IS6 family transposase
					STACADC2_RS00665	pseudo
					STACADC2_RS00670	poly-gamma-glutamate biosynthesis protein
					STACADC2_RS00675	ACT domain-containing protein
2	4,750	653,694	658,444	30.7	STACADC2_RS03525	DUF4062 domain-containing protein
					STACADC2_RS03530	restriction endonuclease subunit S
					STACADC2_RS03535	pseudo
					STACADC2_RS03540	XRE family transcriptional regulator
3	12,475	758,834	771,309	38.9	STACADC2_RS04055	copper-translocating P-type ATPase
					STACADC2_RS04060	transposase
					STACADC2_RS04065	type I restriction endonuclease subunit R
					STACADC2_RS04070	hypothetical protein
					STACADC2_RS04075	cysteine synthase family protein
					STACADC2_RS04080	aminotransferase class V-fold PLP-dependent enzyme
					STACADC2_RS04085	pseudo
					STACADC2_RS04090	pseudo
					STACADC2_RS04095	pseudo
					STACADC2_RS04100	restriction endonuclease subunit S
					STACADC2_RS04105	DNA methyltransferase
					STACADC2_RS04110	pseudo
					STACADC2_RS04115	sugar transporter
4	4,512	776,163	780,675	36.7	STACADC2_RS04150	hypothetical protein
					STACADC2_RS04155	pseudo
					STACADC2_RS04160	putative sulfate exporter family transporter
					STACADC2_RS04165	pseudo
					STACADC2_RS04170	pseudo
5	6,289	818,121	824,410	30.1	STACADC2_RS04360	pseudo
					STACADC2_RS04365	hypothetical protein
					STACADC2_RS04370	XRE family transcriptional regulator
					STACADC2_RS04375	MutR family transcriptional regulator
					STACADC2_RS04380	Fe-S oxidoreductase

					STACADC2_RS04385	radical SAM protein
					STACADC2_RS04390	agmatinase
					STACADC2_RS04395	MFS transporter
6	7,625	931,514	939,139	32.5	STACADC2_RS04995	ISL3-like element ISSth1 family transposase
					STACADC2_RS05000	hypothetical protein
					STACADC2_RS05005	hypothetical protein
					STACADC2_RS05010	hypothetical protein
					STACADC2_RS05015	hypothetical protein
					STACADC2_RS05020	hypothetical protein
					STACADC2_RS05025	hypothetical protein
					STACADC2_RS05030	glycosyltransferase family 2 protein
7	4,183	957,404	961,587	37.7	STACADC2_RS05135	N-acetyltransferase
					STACADC2_RS05140	IS200/IS605 family transposase
					STACADC2_RS05145	50S ribosomal protein L20
					STACADC2_RS05150	50S ribosomal protein L35
					STACADC2_RS05155	translation initiation factor IF-3
					STACADC2_RS05160	cytidylate kinase
					STACADC2_RS05165	LysM domain-containing protein
					STACADC2_RS05170	ferredoxin
					STACADC2_RS05175	hypothetical protein
8	4,150	1,249,165	1,253,315	32.0	STACADC2_RS06635	DUF2142 domain-containing protein
					STACADC2_RS06640	glycosyltransferase family 2 protein
					STACADC2_RS06645	polysaccharide biosynthesis protein
					STACADC2_RS06650	DUF2304 domain-containing protein
9	13,725	1,422,231	1,435,956	38.3	STACADC2_RS07570	thioredoxin
					STACADC2_RS07575	IS3-like element ISSth1b family transposase
					STACADC2_RS07580	ComC/BlpC family peptide pheromone/bacteriocin
					STACADC2_RS07585	hypothetical protein
					STACADC2_RS07590	hypothetical protein
					STACADC2_RS07595	ComC/BlpC family peptide pheromone/bacteriocin
					STACADC2_RS07600	DNA-binding response regulator
					STACADC2_RS07605	ATP-binding protein
					STACADC2_RS07610	ComC/BlpC family peptide pheromone/bacteriocin
					STACADC2_RS07615	pseudo
					STACADC2_RS07620	peptide cleavage/export ABC transporter
					STACADC2_RS07625	hypothetical protein
					STACADC2_RS07630	ABC transporter ATP-binding protein
					STACADC2_RS07635	multidrug ABC transporter permease/ATP-binding protein
10	10,847	1,657,993	1,668,840	31.0	STACADC2_RS08835	helix-turn-helix domain-containing protein
					STACADC2_RS08840	pseudo
					STACADC2_RS08845	pseudo
					STACADC2_RS08850	MutR family transcriptional regulator
					STACADC2_RS08855	pseudo
					STACADC2_RS08860	ABC transporter ATP-binding protein
					STACADC2_RS08865	replication initiation protein
					STACADC2_RS08870	DUF3173 domain-containing protein

Strain APC151						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	11,303	5,515	16,818	37.9	B1761_RS00025	cysteine synthase family protein
					B1761_RS00030	cystathionine gamma-synthase
					B1761_RS00035	serine acetyltransferase
					B1761_RS00040	pseudo
					B1761_RS00045	pseudo
					B1761_RS00050	pseudo
					B1761_RS00055	IS256 family transposase
					B1761_RS00060	sugar transporter
					B1761_RS00065	pseudo
					B1761_RS00070	hypothetical protein
					B1761_RS00075	pseudo
					B1761_RS00080	PrsW family intramembrane metalloprotease
					B1761_RS00085	zinc ABC transporter substrate-binding protein AdcA
					B1761_RS00090	hypothetical protein
2	5,144	52,462	57,606	29.4	B1761_RS00290	MutR family transcriptional regulator
					B1761_RS00295	Fe-S oxidoreductase
					B1761_RS00300	radical SAM protein
					B1761_RS00305	agmatinase
					B1761_RS00310	MFS transporter
3	17,606	163,250	180,856	34.2	B1761_RS00920	AI-2E family transporter
					B1761_RS00925	pseudo
					B1761_RS00930	pseudo
					B1761_RS00935	pseudo
					B1761_RS00940	phosphatase
					B1761_RS00945	VanZ family protein
					B1761_RS00950	hypothetical protein
					B1761_RS00955	pseudo
					B1761_RS00960	glycosyltransferase family 2 protein
					B1761_RS00965	glycosyltransferase family 2 protein
					B1761_RS00970	hypothetical protein
					B1761_RS00975	glycosyl transferase
					B1761_RS00980	multidrug MFS transporter
					B1761_RS00985	UDP-N-acetylglucosamine--LPS N-acetylglucosamine transferase
					B1761_RS00990	pseudo
					B1761_RS00995	IS3 family transposase
					B1761_RS01000	pseudo
					B1761_RS01005	hypothetical protein
					B1761_RS01010	pseudo
					B1761_RS01015	flippase
					B1761_RS01020	pseudo
					B1761_RS01025	glycosyltransferase family 2 protein
					B1761_RS01030	DUF1792 domain-containing protein
					B1761_RS01035	DUF1792 domain-containing protein
4	5,091	451,357	456,448	31.4	B1761_RS02480	transporter
					B1761_RS02485	KxxxW cyclic peptide radical SAM maturase
					B1761_RS02490	KxxxW-cyclized peptide pheromone
					B1761_RS02495	helix-turn-helix domain-containing protein

					B1761_RS02500	pseudo
5	14,915	933,916	948,831	31.4	B1761_RS04905	hypothetical protein
					B1761_RS04910	hypothetical protein
					B1761_RS04915	pseudo
					B1761_RS04920	ATP-binding protein
					B1761_RS04925	dicarboxylate/amino acid:cation symporter
					B1761_RS04930	alpha-L-glutamate ligase
					B1761_RS04935	hypothetical protein
					B1761_RS04940	hypothetical protein
					B1761_RS04945	hypothetical protein
					B1761_RS04950	hypothetical protein
					B1761_RS04955	hypothetical protein
					B1761_RS04960	DUF2513 domain-containing protein
					B1761_RS04965	hypothetical protein
					B1761_RS04970	hypothetical protein
					B1761_RS04975	endonuclease
					B1761_RS04980	XRE family transcriptional regulator
					B1761_RS04985	DNA cytosine methyltransferase
					B1761_RS04990	hypothetical protein
6	16,813	951,824	968,637	39.8	B1761_RS05020	helix-turn-helix domain-containing protein
					B1761_RS05025	IS30 family transposase
					B1761_RS05030	50S ribosomal protein L17
					B1761_RS05035	DNA-directed RNA polymerase subunit alpha
					B1761_RS05040	30S ribosomal protein S11
					B1761_RS05045	30S ribosomal protein S13
					B1761_RS05050	50S ribosomal protein L36
					B1761_RS05055	translation initiation factor IF-1
					B1761_RS05060	adenylate kinase
					B1761_RS05065	preprotein translocase subunit SecY
					B1761_RS05070	50S ribosomal protein L15
					B1761_RS05075	50S ribosomal protein L30
					B1761_RS05080	30S ribosomal protein S5
					B1761_RS05085	50S ribosomal protein L18
					B1761_RS05090	50S ribosomal protein L6
					B1761_RS05095	30S ribosomal protein S8
					rpsN	30S ribosomal protein S14 type Z
					B1761_RS05105	50S ribosomal protein L5
					B1761_RS05110	50S ribosomal protein L24
					B1761_RS05115	50S ribosomal protein L14
					B1761_RS05120	30S ribosomal protein S17
					B1761_RS05125	50S ribosomal protein L29
					B1761_RS05130	50S ribosomal protein L16
					B1761_RS05135	30S ribosomal protein S3
					B1761_RS05140	50S ribosomal protein L22
					B1761_RS05145	30S ribosomal protein S19
					B1761_RS05150	50S ribosomal protein L2
					B1761_RS05155	50S ribosomal protein L23
					B1761_RS05160	50S ribosomal protein L4
					B1761_RS05165	50S ribosomal protein L3
					B1761_RS05170	30S ribosomal protein S10
					B1761_RS05175	Holliday junction branch migration DNA helicase RuvB

7	7,580	973,209	980,789	34.9	B1761_RS05200	adenylate kinase
					B1761_RS05205	MutR family transcriptional regulator
					B1761_RS05210	radical SAM protein
					B1761_RS05215	ABC transporter ATP-binding protein
					B1761_RS05220	XRE family transcriptional regulator
					B1761_RS05225	hypothetical protein
					B1761_RS05230	replication initiator protein
					B1761_RS05235	DUF3173 domain-containing protein
8	11,868	1,150,322	1,162,190	32.2	B1761_RS06220	hypothetical protein
					B1761_RS06225	ketopantoate reductase family protein
					B1761_RS06230	50S ribosomal protein L13
					B1761_RS06235	30S ribosomal protein S9
					B1761_RS06240	pseudo
					B1761_RS06245	helix-turn-helix transcriptional regulator
					B1761_RS06250	hypothetical protein
					B1761_RS06255	pseudo
					B1761_RS06260	lantibiotic biosynthesis protein
					B1761_RS06265	pseudo
					B1761_RS06270	pseudo
					B1761_RS06275	integrase
					B1761_RS06280	Fis family transcriptional regulator
9	22,592	1,416,711	1,439,303	38.4	B1761_RS07695	acyl carrier protein
					B1761_RS07700	enoyl-[acyl-carrier-protein] reductase FabK
					B1761_RS07705	[acyl-carrier-protein] S-malonyltransferase
					B1761_RS07710	3-oxoacyl-[acyl-carrier-protein] reductase
					B1761_RS07715	beta-ketoacyl-[acyl-carrier-protein] synthase II
					B1761_RS07720	acetyl-CoA carboxylase biotin carboxyl carrier protein
					B1761_RS07725	beta-hydroxyacyl-ACP dehydratase
					B1761_RS07730	acetyl-CoA carboxylase biotin carboxylase subunit
					B1761_RS07735	acetyl-CoA carboxylase carboxyl transferase subunit beta
					B1761_RS07740	acetyl-CoA carboxylase carboxyl transferase subunit alpha
					B1761_RS07745	S-ribosylhomocysteine lyase
					B1761_RS07750	pseudo
					B1761_RS07755	type I restriction endonuclease
					B1761_RS07760	pseudo
					B1761_RS07765	glutamate decarboxylase
					B1761_RS07770	amino acid permease
					B1761_RS07775	hypothetical protein
					B1761_RS07780	IS6 family transposase
					B1761_RS07785	hypothetical protein
					B1761_RS07790	DUF3114 domain-containing protein
					B1761_RS07795	pseudo
					B1761_RS07800	ABC transporter ATP-binding protein
					B1761_RS07805	pseudo
					B1761_RS07810	helix-turn-helix domain-containing protein
					B1761_RS07815	pseudo
					B1761_RS07820	IS3 family transposase
10	9,929	1,785,887	1,795,816	35.9	B1761_RS09595	DNA-binding protein
					B1761_RS09600	hypothetical protein
					B1761_RS09605	hypothetical protein
					B1761_RS09610	hypothetical protein

					B1761_RS09615	DNA primase
					B1761_RS09620	hypothetical protein
					B1761_RS09625	hypothetical protein
					B1761_RS09630	hypothetical protein
					B1761_RS09635	hypothetical protein
					B1761_RS09640	hypothetical protein
					B1761_RS09645	hypothetical protein
					B1761_RS09650	XRE family transcriptional regulator
					B1761_RS09655	XRE family transcriptional regulator
					B1761_RS09660	site-specific integrase
					B1761_RS09665	hypothetical protein

Strain ASCC 1275						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	35,727	89,866	125,593	36.2	T303_RS09825	helix-turn-helix domain-containing protein
					T303_RS00505	IS30 family transposase
					T303_RS00510	50S ribosomal protein L17
					T303_RS00515	DNA-directed RNA polymerase subunit alpha
					T303_RS00520	30S ribosomal protein S11
					T303_RS00525	30S ribosomal protein S13
					T303_RS09830	50S ribosomal protein L36
					T303_RS00530	translation initiation factor IF-1
					T303_RS00535	adenylate kinase
					T303_RS00540	preprotein translocase subunit SecY
					T303_RS00545	50S ribosomal protein L15
					T303_RS00550	50S ribosomal protein L30
					T303_RS00555	30S ribosomal protein S5
					T303_RS00560	50S ribosomal protein L18
					T303_RS00565	50S ribosomal protein L6
					T303_RS00570	30S ribosomal protein S8
					rpsN	30S ribosomal protein S14 type Z
					T303_RS00580	50S ribosomal protein L5
					T303_RS00585	50S ribosomal protein L24
					T303_RS00590	50S ribosomal protein L14
					T303_RS00595	30S ribosomal protein S17
					T303_RS00600	50S ribosomal protein L29
					T303_RS00605	50S ribosomal protein L16
					T303_RS00610	30S ribosomal protein S3
					T303_RS00615	50S ribosomal protein L22
					T303_RS00620	30S ribosomal protein S19
					T303_RS00625	50S ribosomal protein L2
					T303_RS00630	50S ribosomal protein L23
					T303_RS00635	50S ribosomal protein L4
					T303_RS00640	50S ribosomal protein L3
					T303_RS00645	30S ribosomal protein S10
					T303_RS00650	Holliday junction branch migration DNA helicase RuvB
					T303_RS00655	acyltransferase
					T303_RS00660	membrane protein
					T303_RS00665	low molecular weight phosphotyrosine protein phosphatase

					T303_RS00670	adenylosuccinate synthetase
					T303_RS00675	pseudo
					T303_RS00680	MFS transporter
					T303_RS00685	bacteriocin biosynthesis protein
					T303_RS00690	IS256 family transposase
					T303_RS00695	hypothetical protein
					T303_RS00700	ABC transporter ATP-binding protein
					T303_RS00705	MutR family transcriptional regulator
					T303_RS00710	radical SAM protein
					T303_RS00715	ABC transporter ATP-binding protein
					T303_RS00720	pseudo
					T303_RS09835	hypothetical protein
					T303_RS09840	replication initiator protein
					T303_RS00735	DUF3173 domain-containing protein
2	25,790	206,528	232,318	42.1	T303_RS01170	transposase
					T303_RS01175	IS3 family transposase
					T303_RS01285	rod shape-determining protein MreC
					T303_RS01290	rod shape-determining protein MreD
					T303_RS01295	CHAP domain-containing protein
					T303_RS01300	ribose-phosphate diphosphokinase
					T303_RS09885	pseudo
					T303_RS01310	pyridoxal phosphate-dependent aminotransferase
					T303_RS01315	DNA repair protein RecO
					T303_RS01320	phosphate acyltransferase PlsX
					T303_RS01325	acyl carrier protein
					T303_RS01330	phosphoribosylaminoimidazolesuccinocarboxamide synthase
					T303_RS01335	phosphoribosylformylglycinamidine synthase
					T303_RS01340	amidophosphoribosyltransferase
					T303_RS01345	phosphoribosylformylglycinamidine cyclo-ligase
					T303_RS01350	phosphoribosylglycinamide formyltransferase
					T303_RS01355	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/inosine monophosphate cyclohy
3	13,289	289,525	302,814	33.0	T303_RS01700	hypothetical protein
					T303_RS01705	2-dehydropantoate 2-reductase
					T303_RS01710	50S ribosomal protein L13
					T303_RS01715	30S ribosomal protein S9
					T303_RS01720	ISL3 family transposase
					T303_RS09915	pseudo
					T303_RS01725	helix-turn-helix transcriptional regulator
					T303_RS01730	hypothetical protein
					T303_RS01735	pseudo
					T303_RS01740	lantibiotic biosynthesis protein
					T303_RS01745	lantibiotic transporter
					T303_RS01750	pseudo
					T303_RS01755	pseudo
					T303_RS01760	DNA-binding protein
4	7,495	679,551	687,046	35.8	T303_RS03735	ABC transporter substrate-binding protein
					T303_RS10080	hypothetical protein
					T303_RS10085	pseudo
					T303_RS03745	macrolide ABC transporter
					T303_RS03750	ABC transporter permease, truncated

					T303_RS03755	50S ribosomal protein L10
					T303_RS03760	50S ribosomal protein L7/L12
					T303_RS03770	pseudo
					T303_RS10090	hypothetical protein
					T303_RS03775	ABC transporter ATP-binding protein
					T303_RS03780	ABC transporter permease
					T303_RS10095	hypothetical protein
					T303_RS03785	hypothetical protein
5	5,578	861,889	867,467	29.3	T303_RS04685	hypothetical protein
					T303_RS04690	hypothetical protein
					T303_RS04695	hypothetical protein
					T303_RS04700	hypothetical protein
					T303_RS04705	restriction endonuclease subunit S
					T303_RS04710	NTPase
6	7,212	927,283	934,495	37.0	T303_RS05010	hypothetical protein
					T303_RS05015	hypothetical protein
					T303_RS05020	DNA primase
					T303_RS05025	hypothetical protein
					T303_RS05030	hypothetical protein
					T303_RS10200	pseudo
					T303_RS05040	hypothetical protein
					T303_RS05045	hypothetical protein
					T303_RS05050	hypothetical protein
					T303_RS05055	XRE family transcriptional regulator
					T303_RS10205	XRE family transcriptional regulator
					T303_RS05065	site-specific integrase
					T303_RS05070	hypothetical protein
7	21,284	987,436	1,008,720	37.5	T303_RS05340	cold-shock protein
					T303_RS05345	cold-shock protein
					T303_RS10255	pseudo
					T303_RS05360	IS256 family transposase
					T303_RS05365	pseudo
					T303_RS05370	copper-translocating P-type ATPase
					T303_RS05375	pseudo
					T303_RS05380	type I restriction endonuclease subunit R
					T303_RS10260	pseudo
					T303_RS05385	cysteine synthase family protein
					T303_RS05390	cystathionine gamma-synthase
					T303_RS10265	pseudo
					T303_RS10270	pseudo
					T303_RS05400	pseudo
					T303_RS10275	pseudo
					T303_RS05415	DNA methyltransferase
					T303_RS10280	pseudo
					T303_RS05420	sugar transporter
					T303_RS10285	pseudo
					T303_RS05425	hypothetical protein
					T303_RS05430	pseudo
					T303_RS05435	PrsW family intramembrane metalloprotease
					T303_RS05440	zinc ABC transporter substrate-binding protein AdcA
					T303_RS05445	hypothetical protein

8	6,465	1,044,658	1,051,123	31.2	T303_RS05640	MutR family transcriptional regulator
					T303_RS05645	Fe-S oxidoreductase
					T303_RS05650	IS256 family transposase
					T303_RS05655	radical SAM protein
					T303_RS05660	pseudo
					T303_RS05665	MFS transporter
9	13,041	1,142,208	1,155,249	36.0	T303_RS06135	pseudo
					T303_RS06140	pseudo
					T303_RS06145	type I-E CRISPR-associated endoribonuclease Cas2
					T303_RS06150	type I-E CRISPR-associated endonuclease Cas1
					T303_RS06155	type I-E CRISPR-associated protein Cas6/Cse3/CasE
					T303_RS06160	type I-E CRISPR-associated protein Cas5/CasD
					T303_RS06165	type I-E CRISPR-associated protein Cas7/Cse4/CasC
					T303_RS06170	type I-E CRISPR-associated protein Cse2/CasB
					T303_RS06175	pseudo
					T303_RS06180	CRISPR-associated helicase/endonuclease Cas3
					T303_RS06185	DUF2207 domain-containing protein
10	5,605	1,391,298	1,396,903	26.4	T303_RS07420	MFS transporter permease
					T303_RS07425	pseudo
					T303_RS07430	helix-turn-helix domain-containing protein
					T303_RS07435	NUDIX hydrolase
					T303_RS07440	hypothetical protein
11	7,681	1,437,339	1,445,020	30.8	T303_RS07665	ABC transporter ATP-binding protein
					T303_RS07670	hypothetical protein
					T303_RS07675	hypothetical protein
					T303_RS07680	sodium transporter
					T303_RS07685	pseudo
					T303_RS07690	SPASM domain-containing protein
					T303_RS07695	transporter
					T303_RS07700	pseudo
					T303_RS10520	KxxxW-cyclized peptide pheromone
					T303_RS07705	helix-turn-helix domain-containing protein

Strain B59671						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	6,287	164,494	170,781	30.2	CG712_RS00920	pseudo
					CG712_RS00925	hypothetical protein
					CG712_RS00930	XRE family transcriptional regulator
					CG712_RS00935	MutR family transcriptional regulator
					CG712_RS00940	Fe-S oxidoreductase
					CG712_RS00945	radical SAM protein
					CG712_RS00950	agmatinase
					CG712_RS00955	MFS transporter
2	4,445	324,830	329,275	33.6	CG712_RS01705	hypothetical protein
					CG712_RS01710	hypothetical protein
					CG712_RS01715	pseudo
					CG712_RS01720	hypothetical protein
					CG712_RS01725	hypothetical protein
					CG712_RS01730	hypothetical protein

3	4,173	560,018	564,191	40.6	CG712_RS02935	uroporphyrinogen decarboxylase
					CG712_RS02940	glutathione peroxidase
					CG712_RS02945	hypothetical protein
					CG712_RS02950	pseudo
					CG712_RS02955	pseudo
					CG712_RS02960	TetR/AcrR family transcriptional regulator
					CG712_RS02965	pseudo
4	10,849	1,075,104	1,085,953	31.0	CG712_RS05625	helix-turn-helix domain-containing protein
					CG712_RS05630	hypothetical protein
					CG712_RS05635	ABC transporter ATP-binding protein
					CG712_RS05640	MutR family transcriptional regulator
					CG712_RS05645	pseudo
					CG712_RS05650	ABC transporter ATP-binding protein
					CG712_RS05655	pseudo
					CG712_RS05660	DUF3173 domain-containing protein
5	10,973	1,799,394	1,810,367	40.4	CG712_RS09525	ComC/BlpC family peptide pheromone/bacteriocin
					CG712_RS09530	pseudo
					CG712_RS09535	hypothetical protein
					CG712_RS09540	plasmid mobilization relaxosome protein MobC
					CG712_RS09545	relaxase
					CG712_RS09550	hypothetical protein
					CG712_RS09555	LytTR family transcriptional regulator
					CG712_RS09560	DNA-binding response regulator
					CG712_RS09565	GHKL domain-containing protein
					CG712_RS09570	ComC/BlpC family peptide pheromone/bacteriocin
					CG712_RS09575	bacteriocin secretion accessory protein
					CG712_RS09580	peptide cleavage/export ABC transporter
					CG712_RS09585	ThmA bacteriocin
					CG712_RS09590	class IIb bacteriocin, lactobin A/cerein 7B family

Strain CNRZ1066						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	14,523	99,677	114,200	32.5	STR_RS00605	hypothetical protein
					STR_RS00610	2-dehydropantoate 2-reductase
					STR_RS00615	50S ribosomal protein L13
					STR_RS00620	30S ribosomal protein S9
					STR_RS09610	hypothetical protein
					STR_RS09615	pseudo
					STR_RS00625	pseudo
					STR_RS00630	hypothetical protein
					STR_RS00635	pseudo
					STR_RS00640	lantibiotic biosynthesis protein
					STR_RS00645	lantibiotic transporter
					STR_RS00650	pseudo
					STR_RS00655	pseudo
					STR_RS00660	pseudo
					STR_RS00665	pseudo
					STR_RS00670	IS30 family transposase
2	17,365	642,151	659,516	35.5	STR_RS03470	(deoxy)nucleoside triphosphate pyrophosphohydrolase

					STR_RS03475	pseudo
					STR_RS09820	pseudo
					STR_RS03480	hypothetical protein
					STR_RS03485	hypothetical protein
					STR_RS03490	pseudo
					STR_RS03495	hypothetical protein
					STR_RS03500	hypothetical protein
					STR_RS03505	hypothetical protein
					STR_RS03510	XRE family transcriptional regulator
					STR_RS03515	hypothetical protein
					STR_RS03520	IS30 family transposase
					STR_RS03525	lysine--tRNA ligase
					STR_RS03530	LysM domain-containing protein
					STR_RS03535	hypothetical protein
					STR_RS03540	hypothetical protein
					STR_RS03545	HAD family hydrolase
					STR_RS03550	hypothetical protein
					STR_RS03555	histidine phosphatase family protein
					STR_RS03560	aminoacyl-tRNA deacylase
					STR_RS03565	N-acetylmuramoyl-L-alanine amidase
					STR_RS03570	hypothetical protein
					STR_RS03575	hypothetical protein
					STR_RS03580	DUF2974 domain-containing protein
					STR_RS09825	DUF2974 domain-containing protein
					STR_RS09830	hypothetical protein
					STR_RS03585	DUF2974 domain-containing protein
3	5,627	663,869	669,496	30.2	STR_RS03600	ATP-binding protein
					STR_RS03605	hypothetical protein
					STR_RS03610	restriction endonuclease subunit S
					STR_RS03615	NTPase
4	19,491	776,536	796,027	37.4	STR_RS04205	cold-shock protein
					STR_RS04210	cold-shock protein
					STR_RS04215	IS256 family transposase
					STR_RS04220	copper-translocating P-type ATPase
					STR_RS04225	pseudo
					STR_RS04230	type I restriction endonuclease subunit R
					STR_RS09890	hypothetical protein
					STR_RS04235	cysteine synthase family protein
					STR_RS04240	cystathionine gamma-synthase
					STR_RS04245	pseudo
					STR_RS09895	pseudo
					STR_RS04250	IS256 family transposase
					STR_RS09900	pseudo
					STR_RS09905	pseudo
					STR_RS04265	pseudo
					STR_RS04270	pseudo
					STR_RS04275	sugar transporter
					STR_RS09910	pseudo
					STR_RS04280	type I restriction-modification system subunit R
					STR_RS04285	pseudo
					STR_RS09915	pseudo

					STR_RS04295	zinc ABC transporter substrate-binding protein AdcA
					STR_RS04300	hypothetical protein
5	6,166	839,629	845,795	29.9	STR_RS04520	pseudo
					STR_RS04525	XRE family transcriptional regulator
					STR_RS04530	MutR family transcriptional regulator
					STR_RS04535	Fe-S oxidoreductase
					STR_RS04540	radical SAM protein
					STR_RS04545	agmatinase
					STR_RS04550	MFS transporter
6	6,935	933,008	939,943	30.3	STR_RS05025	IS30 family transposase
					STR_RS05030	hypothetical protein
					STR_RS05035	UDP-N-acetylglucosamine enolpyruvyl transferase
					STR_RS05040	NUDIX domain-containing protein
					STR_RS05045	tyrosine--tRNA ligase
					STR_RS05050	MutR family transcriptional regulator
7	12,719	960,253	972,972	34.7	STR_RS05165	AI-2E family transporter
					STR_RS05170	pseudo
					STR_RS05175	pseudo
					STR_RS05180	IS256 family transposase
					STR_RS05185	lipopolysaccharide 1,6-galactosyltransferase
					STR_RS05190	pseudo
					STR_RS05195	VanZ family protein
					STR_RS05200	exopolysaccharide biosynthesis protein
					STR_RS05205	exopolysaccharide polymerization protein
					STR_RS05210	exopolysaccharide gene cluster protein
					STR_RS05215	EpsG family protein
					STR_RS05220	glycosyltransferase family 2 protein
					STR_RS05225	acyltransferase
8	4,087	1,191,550	1,195,637	35.8	STR_RS06365	amino acid ABC transporter permease
					STR_RS06370	GNAT family N-acetyltransferase
					STR_RS06375	hypothetical protein
					STR_RS06380	pseudo
					STR_RS06385	glycerol dehydrogenase
					STR_RS06390	glycerol dehydrogenase
					STR_RS06395	hypothetical protein
9	4,591	1,336,631	1,341,222	32.8	STR_RS07050	hypothetical protein
					STR_RS07055	hypothetical protein
					STR_RS07060	transcriptional regulator
					STR_RS07065	hypothetical protein
					STR_RS07070	helix-turn-helix domain-containing protein
					STR_RS07075	radical SAM/SPASM domain-containing protein
10	25,706	1,475,470	1,501,176	39.7	STR_RS07785	hypothetical protein
					STR_RS07790	DNA polymerase IV
					STR_RS07795	formate C-acetyltransferase
					STR_RS07800	carbonate dehydratase
					STR_RS07805	pseudo
					STR_RS10250	pseudo
					STR_RS07810	type I restriction-modification system subunit S
					STR_RS07815	plasmid mobilization protein
					STR_RS07820	recombinase
					STR_RS07825	DNA recombinase

					STR_RS07830	hypothetical protein
					STR_RS07835	iron export ABC transporter permease subunit FetB
					STR_RS07840	putrescine/spermidine ABC transporter ATP-binding protein
					STR_RS07845	hypothetical protein
					STR_RS07850	serine hydrolase
					STR_RS07855	peptidase
					STR_RS07860	IS30 family transposase
					STR_RS07865	CPBP family intramembrane metalloprotease
					STR_RS07870	aquaporin family protein
					STR_RS07875	Xaa-Pro dipeptidyl-peptidase
					STR_RS07880	hypothetical protein
					STR_RS10255	pseudo
					STR_RS07895	ComC/BlpC family peptide pheromone/bacteriocin
					STR_RS07900	DNA-binding response regulator
					STR_RS07905	ATP-binding protein
					STR_RS07910	ComC/BlpC family peptide pheromone/bacteriocin
					STR_RS07915	pseudo
					STR_RS07920	peptide cleavage/export ABC transporter
11	15,951	1,717,663	1,733,614	30.9	STR_RS09110	DNA topology modulation protein FlaR
					STR_RS09115	MFS transporter
					STR_RS09120	bacteriocin biosynthesis protein
					STR_RS09125	pseudo
					STR_RS09130	helix-turn-helix domain-containing protein
					STR_RS09135	hypothetical protein
					STR_RS09140	pseudo
					STR_RS09145	MutR family transcriptional regulator
					STR_RS09150	radical SAM protein
					STR_RS09155	ABC transporter ATP-binding protein
					STR_RS09165	replication initiation protein
					STR_RS09170	DUF3173 domain-containing protein

Strain CS8						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	6,166	69,845	76,011	29.9	BAY21_RS00380	MFS transporter
					BAY21_RS00385	agmatinase
					BAY21_RS00390	radical SAM protein
					BAY21_RS00395	Fe-S oxidoreductase
					BAY21_RS00400	MutR family transcriptional regulator
					BAY21_RS00405	XRE family transcriptional regulator
					BAY21_RS00410	pseudo
2	12,538	124,260	136,798	38.8	BAY21_RS00660	sugar transporter
					BAY21_RS00665	pseudo
					BAY21_RS00670	pseudo
					BAY21_RS09520	pseudo
					BAY21_RS09525	pseudo
					BAY21_RS00685	IS256 family transposase
					BAY21_RS09530	pseudo
					BAY21_RS00690	pseudo
					BAY21_RS00695	cystathionine gamma-synthase

					BAY21_RS00700	cysteine synthase family protein
					BAY21_RS09535	hypothetical protein
					BAY21_RS00705	type I restriction endonuclease subunit R
					BAY21_RS00710	IS256 family transposase
					BAY21_RS00715	copper-translocating P-type ATPase
3	7,605	179,205	186,810	37.5	BAY21_RS00960	site-specific integrase
					BAY21_RS00965	XRE family transcriptional regulator
					BAY21_RS00970	XRE family transcriptional regulator
					BAY21_RS00975	hypothetical protein
					BAY21_RS00980	hypothetical protein
					BAY21_RS00985	hypothetical protein
					BAY21_RS00990	hypothetical protein
					BAY21_RS00995	hypothetical protein
					BAY21_RS01000	hypothetical protein
					BAY21_RS01005	DNA primase
					BAY21_RS01010	hypothetical protein
					BAY21_RS01015	hypothetical protein
					BAY21_RS01020	hypothetical protein
4	5,627	246,145	251,772	30.2	BAY21_RS01305	NTPase
					BAY21_RS01310	restriction endonuclease subunit S
					BAY21_RS01315	hypothetical protein
					BAY21_RS01320	ATP-binding protein
5	4,411	267,365	271,776	26.5	BAY21_RS01400	hypothetical protein
					BAY21_RS01405	XRE family transcriptional regulator
					BAY21_RS01410	hypothetical protein
					BAY21_RS01415	hypothetical protein
					BAY21_RS01420	hypothetical protein
					BAY21_RS01425	hypothetical protein
					BAY21_RS01430	hypothetical protein
					BAY21_RS01435	hypothetical protein
6	19,736	792,107	811,843	34.0	BAY21_RS04145	ACT domain-containing protein
					BAY21_RS04150	hypothetical protein
					BAY21_RS04155	pseudo
					BAY21_RS04165	pseudo
					BAY21_RS04170	pseudo
					BAY21_RS04175	IS30 family transposase
					BAY21_RS04180	pseudo
					BAY21_RS04185	pseudo
					BAY21_RS04190	pseudo
					BAY21_RS04195	pseudo
					BAY21_RS04200	lantibiotic transporter
					BAY21_RS04205	lantibiotic biosynthesis protein
					BAY21_RS04210	pseudo
					BAY21_RS04215	hypothetical protein
					BAY21_RS04220	pseudo
					BAY21_RS09805	pseudo
					BAY21_RS09810	hypothetical protein
					BAY21_RS04225	30S ribosomal protein S9
					BAY21_RS04230	50S ribosomal protein L13
					BAY21_RS04235	2-dehydropantoate 2-reductase
					BAY21_RS04240	hypothetical protein

					BAY21_RS09815	transcriptional regulator
					BAY21_RS04245	pseudo
					BAY21_RS04250	hypothetical protein
7	15,951	972,447	988,398	30.9	BAY21_RS05185	DUF3173 domain-containing protein
					BAY21_RS05190	replication initiation protein
					BAY21_RS05200	ABC transporter ATP-binding protein
					BAY21_RS05205	radical SAM protein
					BAY21_RS05210	MutR family transcriptional regulator
					BAY21_RS05215	pseudo
					BAY21_RS05220	hypothetical protein
					BAY21_RS05225	helix-turn-helix domain-containing protein
					BAY21_RS05230	pseudo
					BAY21_RS05235	bacteriocin biosynthesis protein
					BAY21_RS05240	MFS transporter
					BAY21_RS05245	DNA topology modulation protein FlaR
8	21,101	1,204,204	1,225,305	39.4	BAY21_RS09960	hypothetical protein
					BAY21_RS06450	peptide cleavage/export ABC transporter
					BAY21_RS06455	pseudo
					BAY21_RS06460	ComC/BlpC family peptide pheromone/bacteriocin
					BAY21_RS06465	ATP-binding protein
					BAY21_RS06470	DNA-binding response regulator
					BAY21_RS06475	ComC/BlpC family peptide pheromone/bacteriocin
					BAY21_RS09965	pseudo
					BAY21_RS06490	hypothetical protein
					BAY21_RS06495	Xaa-Pro dipeptidyl-peptidase
					BAY21_RS06500	aquaporin family protein
					BAY21_RS06505	CPBP family intramembrane metalloprotease
					BAY21_RS06510	IS30 family transposase
					BAY21_RS06515	peptidase
					BAY21_RS06520	serine hydrolase
					BAY21_RS06525	hypothetical protein
					BAY21_RS06530	ABC transporter ATPase
					BAY21_RS06535	iron export ABC transporter permease subunit FetB
					BAY21_RS06540	hypothetical protein
					BAY21_RS06545	DNA recombinase
					BAY21_RS06550	recombinase
					BAY21_RS06555	plasmid mobilization protein
					BAY21_RS06560	type I restriction-modification system subunit S
					BAY21_RS09970	pseudo
					BAY21_RS06565	pseudo
9	7,879	1,510,263	1,518,142	35.6	BAY21_RS07960	amidohydrolase
					BAY21_RS07965	hypothetical protein
					BAY21_RS07970	glycerol dehydrogenase
					BAY21_RS07975	glycerol dehydrogenase
					BAY21_RS07980	pseudo
					BAY21_RS07985	hypothetical protein
					BAY21_RS07990	GNAT family N-acetyltransferase
					BAY21_RS07995	amino acid ABC transporter permease
					BAY21_RS08000	amino acid ABC transporter ATP-binding protein
					BAY21_RS08005	adenylosuccinate lyase
					BAY21_RS08010	amino acid ABC transporter substrate-binding protein

10	10,035	1,732,969	1,743,004	33.0	BAY21_RS09125	glycosyltransferase family 1 protein
					BAY21_RS09130	acyltransferase
					BAY21_RS09135	glycosyltransferase family 2 protein
					BAY21_RS09140	EpsG family protein
					BAY21_RS09145	exopolysaccharide gene cluster protein
					BAY21_RS09150	exopolysaccharide polymerization protein
					BAY21_RS09155	exopolysaccharide biosynthesis protein
					BAY21_RS09160	VanZ family protein
					BAY21_RS09165	pseudo
					BAY21_RS09170	lipopolysaccharide 1,6-galactosyltransferase
11	5,850	1,767,354	1,773,204	28.8	BAY21_RS09310	MutR family transcriptional regulator
					BAY21_RS09315	tyrosine--tRNA ligase
					BAY21_RS09320	NUDIX domain-containing protein
					BAY21_RS09325	UDP-N-acetylglucosamine enolpyruvyl transferase
					BAY21_RS09330	hypothetical protein

Strain DGCC 7710						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	12,728	101,144	113,872	32.3	CW339_RS00585	50S ribosomal protein L13
					CW339_RS00590	30S ribosomal protein S9
					CW339_RS00595	ISL3 family transposase
					CW339_RS00600	pseudo
					CW339_RS00605	helix-turn-helix transcriptional regulator
					CW339_RS00610	hypothetical protein
					CW339_RS00615	pseudo
					CW339_RS00620	lantibiotic biosynthesis protein
					CW339_RS00625	lantibiotic transporter
					CW339_RS00630	pseudo
					CW339_RS00635	pseudo
					CW339_RS00640	DNA-binding protein
					CW339_RS00645	hypothetical protein
2	4,715	489,352	494,067	38.7	CW339_RS02625	ABC transporter substrate-binding protein
					CW339_RS02630	hypothetical protein
					CW339_RS02635	pseudo
					CW339_RS02640	macrolide ABC transporter
					CW339_RS02645	ABC transporter permease, truncated
					CW339_RS02650	50S ribosomal protein L10
					CW339_RS02655	50S ribosomal protein L7/L12
					CW339_RS02660	pseudo
					CW339_RS02665	hypothetical protein
3	5,579	675,213	680,792	29.3	CW339_RS03635	hypothetical protein
					CW339_RS03640	hypothetical protein
					CW339_RS03645	hypothetical protein
					CW339_RS03650	restriction endonuclease subunit S
					CW339_RS03655	NTPase
4	7,212	740,609	747,821	37.0	CW339_RS03950	hypothetical protein
					CW339_RS03955	hypothetical protein
					CW339_RS03960	DNA primase
					CW339_RS03965	hypothetical protein

					CW339_RS03970	hypothetical protein
					CW339_RS03975	pseudo
					CW339_RS03980	hypothetical protein
					CW339_RS03985	hypothetical protein
					CW339_RS03990	hypothetical protein
					CW339_RS03995	XRE family transcriptional regulator
					CW339_RS04000	XRE family transcriptional regulator
					CW339_RS04005	site-specific integrase
					CW339_RS04010	hypothetical protein
5	12,439	804,762	817,201	39.0	CW339_RS04295	copper-translocating P-type ATPase
					CW339_RS04300	IS256 family transposase
					CW339_RS04305	pseudo
					CW339_RS04310	pseudo
					CW339_RS04315	cysteine synthase family protein
					CW339_RS04320	aminotransferase class V-fold PLP-dependent enzyme
					CW339_RS04325	pseudo
					CW339_RS04330	XRE family transcriptional regulator
					CW339_RS04335	IS256 family transposase
					CW339_RS04340	pseudo
					CW339_RS04345	DNA methyltransferase
					CW339_RS04350	pseudo
					CW339_RS04355	sugar transporter
6	6,466	857,779	864,245	31.2	CW339_RS04585	Rgg/GadR/MutR family transcriptional regulator
					CW339_RS04590	Fe-S oxidoreductase
					CW339_RS04595	IS256 family transposase
					CW339_RS04600	radical SAM protein
					CW339_RS04605	arginase family protein
					CW339_RS04610	MFS transporter
7	13,042	955,327	968,369	36.0	CW339_RS05105	pseudo
					CW339_RS05110	pseudo
					cas2e	type I-E CRISPR-associated endoribonuclease Cas2
					CW339_RS05120	type I-E CRISPR-associated endonuclease Cas1
					cas6e	type I-E CRISPR-associated protein Cas6/Cse3/CasE
					cas5e	type I-E CRISPR-associated protein Cas5/CasD
					cas7e	type I-E CRISPR-associated protein Cas7/Cse4/CasC
					casB	type I-E CRISPR-associated protein Cse2/CasB
					CW339_RS05145	type I-E CRISPR-associated protein Cse1/CasA
					CW339_RS05150	CRISPR-associated helicase/endonuclease Cas3
					CW339_RS05155	DUF2207 domain-containing protein
8	5,607	1,204,419	1,210,026	26.4	CW339_RS06415	MFS transporter permease
					CW339_RS06420	oligoendopeptidase F
					CW339_RS06425	Rgg/GadR/MutR family transcriptional regulator
					CW339_RS06430	NUDIX hydrolase
					CW339_RS06435	hypothetical protein
9	4,677	1,252,707	1,257,384	30.3	CW339_RS06680	sodium transporter
					CW339_RS06685	pseudo
					CW339_RS06690	SPASM domain-containing protein
					CW339_RS06695	pseudo
					CW339_RS06700	transporter
					kwcM	pseudo
					CW339_RS06710	KxxxW-cyclized peptide pheromone

					CW339_RS06715	Rgg/GadR/MutR family transcriptional regulator
10	8,496	1,525,756	1,534,252	37.5	CW339_RS08065	IS3-like element ISSth1b family transposase
					CW339_RS08070	hypothetical protein
					CW339_RS08075	hypothetical protein
					CW339_RS08080	hypothetical protein
					CW339_RS08085	ComC/BlpC family peptide pheromone/bacteriocin
					CW339_RS08090	hypothetical protein
					CW339_RS08095	ComC/BlpC family peptide pheromone/bacteriocin
					CW339_RS08100	DNA-binding response regulator
					CW339_RS08105	GHKL domain-containing protein
					CW339_RS08110	ComC/BlpC family peptide pheromone/bacteriocin
					CW339_RS08115	bacteriocin secretion accessory protein
11	4,070	1,738,747	1,742,817	38.5	CW339_RS09155	type I restriction-modification system subunit M
					CW339_RS09160	pseudo
					CW339_RS09165	DNA-binding protein
12	39,057	1,747,664	1,786,721	35.6	CW339_RS09195	hypothetical protein
					CW339_RS09200	hypothetical protein
					CW339_RS09205	helix-turn-helix domain-containing protein
					CW339_RS09210	IS30 family transposase
					CW339_RS09215	50S ribosomal protein L17
					CW339_RS09220	DNA-directed RNA polymerase subunit alpha
					CW339_RS09225	30S ribosomal protein S11
					CW339_RS09230	30S ribosomal protein S13
					CW339_RS09235	50S ribosomal protein L36
					CW339_RS09240	translation initiation factor IF-1
					CW339_RS09245	adenylate kinase
					CW339_RS09250	preprotein translocase subunit SecY
					CW339_RS09255	50S ribosomal protein L15
					CW339_RS09260	50S ribosomal protein L30
					CW339_RS09265	30S ribosomal protein S5
					CW339_RS09270	50S ribosomal protein L18
					CW339_RS09275	50S ribosomal protein L6
					CW339_RS09280	30S ribosomal protein S8
					CW339_RS09285	type Z 30S ribosomal protein S14
					CW339_RS09290	50S ribosomal protein L5
					CW339_RS09295	50S ribosomal protein L24
					CW339_RS09300	50S ribosomal protein L14
					CW339_RS09305	30S ribosomal protein S17
					CW339_RS09310	50S ribosomal protein L29
					CW339_RS09315	50S ribosomal protein L16
					CW339_RS09320	30S ribosomal protein S3
					CW339_RS09325	50S ribosomal protein L22
					CW339_RS09330	30S ribosomal protein S19
					CW339_RS09335	50S ribosomal protein L2
					CW339_RS09340	50S ribosomal protein L23
					CW339_RS09345	50S ribosomal protein L4
					CW339_RS09350	50S ribosomal protein L3
					CW339_RS09355	30S ribosomal protein S10
					CW339_RS09360	Holliday junction branch migration DNA helicase RuvB
					CW339_RS09365	acyltransferase
					CW339_RS09370	membrane protein

					CW339_RS09375	low molecular weight phosphotyrosine protein phosphatase
					CW339_RS09380	adenylosuccinate synthetase
					CW339_RS09385	pseudo
					CW339_RS09390	MFS transporter
					CW339_RS09395	ThiF family adenylyltransferase
					CW339_RS09400	pseudo
					CW339_RS09405	Rgg/GadR/MutR family transcriptional regulator
					CW339_RS09410	hypothetical protein
					CW339_RS09415	ABC transporter ATP-binding protein
					CW339_RS09420	MutR family transcriptional regulator
					CW339_RS09425	radical SAM protein
					CW339_RS09430	ABC transporter ATP-binding protein
					CW339_RS09435	pseudo
					CW339_RS09440	pseudo
					CW339_RS09445	DUF3173 domain-containing protein
13	10,748	1,807,929	1,818,677	34.6	CW339_RS09595	alkaline shock response membrane anchor protein AmaP
					CW339_RS09600	DUF2273 domain-containing protein
					CW339_RS09605	Asp23/Gls24 family envelope stress response protein
					CW339_RS09610	CsbD family protein
					CW339_RS09615	phosphohydrolase
					CW339_RS09620	IS256 family transposase
					CW339_RS09625	pseudo
					CW339_RS09630	hypothetical protein
					CW339_RS09635	hypothetical protein
					CW339_RS09640	YhgE/Pip domain-containing protein
					CW339_RS09645	TetR/AcrR family transcriptional regulator
					CW339_RS09650	DUF1304 domain-containing protein
					CW339_RS09655	30S ribosomal protein S4

Strain EPS						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	9,190	120,757	129,947	30.1	CR922_RS00680	pseudo
					CR922_RS00685	lantibiotic biosynthesis protein
					CR922_RS00690	lantibiotic transporter
					CR922_RS00695	pseudo
					CR922_RS00700	pseudo
					CR922_RS00705	pseudo
					CR922_RS00710	pseudo
2	8,605	283,169	291,774	41.0	CR922_RS01495	serine hydrolase
					CR922_RS01500	hypothetical protein
					CR922_RS01505	ABC transporter ATPase
					CR922_RS01510	iron export ABC transporter permease subunit FetB
					CR922_RS01515	hypothetical protein
					CR922_RS01520	pseudo
					CR922_RS01525	recombinase
					CR922_RS01530	plasmid mobilization protein
					CR922_RS01535	type I restriction-modification system subunit S
					CR922_RS01540	pseudo
					CR922_RS01545	pseudo

					CR922_RS01550	pseudo
					pflB	formate C-acetyltransferase
3	7,880	567,225	575,105	35.6	CR922_RS03000	amidohydrolase
					CR922_RS03005	hypothetical protein
					CR922_RS03010	glycerol dehydrogenase
					CR922_RS03015	glycerol dehydrogenase
					CR922_RS03020	aminotransferase
					CR922_RS03025	hypothetical protein
					CR922_RS03030	GNAT family N-acetyltransferase
					CR922_RS03035	amino acid permease
					CR922_RS03040	amino acid ABC transporter ATP-binding protein
					CR922_RS03045	adenylosuccinate lyase
					CR922_RS03050	amino acid ABC transporter substrate-binding protein
4	6,166	920,503	926,669	29.9	CR922_RS04875	MFS transporter
					CR922_RS04880	agmatinase
					CR922_RS04885	radical SAM protein
					CR922_RS04890	Fe-S oxidoreductase
					CR922_RS04895	MutR family transcriptional regulator
					CR922_RS04900	XRE family transcriptional regulator
					CR922_RS04905	pseudo
5	4,333	951,757	956,090	40.3	CR922_RS05030	hypothetical protein
					CR922_RS05035	site-specific DNA-methyltransferase
					CR922_RS05040	type III restriction endonuclease
					CR922_RS05045	pseudo
6	7,183	1,028,252	1,035,435	37.5	CR922_RS05480	site-specific integrase
					CR922_RS05485	XRE family transcriptional regulator
					CR922_RS05490	transcriptional regulator
					CR922_RS05495	hypothetical protein
					CR922_RS05500	hypothetical protein
					CR922_RS05505	hypothetical protein
					CR922_RS05510	hypothetical protein
					CR922_RS05515	hypothetical protein
					CR922_RS05520	hypothetical protein
					CR922_RS05525	DNA primase
					CR922_RS05530	hypothetical protein
					CR922_RS05535	hypothetical protein
7	4,694	1,095,834	1,100,528	28.2	CR922_RS05830	restriction endonuclease subunit S
					CR922_RS05835	pseudo
					CR922_RS05840	hypothetical protein
8	4,411	1,116,163	1,120,574	26.5	CR922_RS05930	hypothetical protein
					CR922_RS05935	XRE family transcriptional regulator
					CR922_RS05940	hypothetical protein
					CR922_RS05945	hypothetical protein
					CR922_RS05950	hypothetical protein
					CR922_RS05955	hypothetical protein
					CR922_RS05960	hypothetical protein
					CR922_RS05965	hypothetical protein
9	15,950	1,733,570	1,749,520	30.9	CR922_RS09270	DNA topology modulation protein FlaR
					CR922_RS09275	MFS transporter
					CR922_RS09280	MccB-like protein
					CR922_RS09285	pseudo

					CR922_RS09290	helix-turn-helix domain-containing protein
					CR922_RS09295	hypothetical protein
					CR922_RS09300	pseudo
					CR922_RS09305	MutR family transcriptional regulator
					CR922_RS09310	radical SAM protein
					CR922_RS09315	ABC transporter ATP-binding protein
					CR922_RS09320	replication initiation protein
					CR922_RS09325	DUF3173 domain-containing protein

Strain GABA						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	11,221	101,513	112,734	32.1	CR921_RS00590	2-dehydropantoate 2-reductase
					CR921_RS00595	50S ribosomal protein L13
					CR921_RS00600	30S ribosomal protein S9
					CR921_RS00605	hypothetical protein
					CR921_RS00610	pseudo
					CR921_RS00615	helix-turn-helix transcriptional regulator
					CR921_RS00620	hypothetical protein
					CR921_RS00625	pseudo
					CR921_RS00630	lantibiotic biosynthesis protein
					CR921_RS00635	lantibiotic transporter
					CR921_RS00640	pseudo
					CR921_RS00645	site-specific integrase
					CR921_RS00650	DNA-binding protein
2	18,227	363,404	381,631	38.7	CR921_RS02020	acyl carrier protein
					CR921_RS02025	enoyl-[acyl-carrier-protein] reductase FabK
					fabD	[acyl-carrier-protein] S-malonyltransferase
					fabG	3-oxoacyl-[acyl-carrier-protein] reductase
					fabF	beta-ketoacyl-[acyl-carrier-protein] synthase II
					CR921_RS02045	acetyl-CoA carboxylase biotin carboxyl carrier protein
					fabZ	beta-hydroxyacyl-ACP dehydratase
					accC	acetyl-CoA carboxylase biotin carboxylase subunit
					CR921_RS02060	acetyl-CoA carboxylase carboxyltransferase subunit beta
					CR921_RS02065	acetyl-CoA carboxylase carboxyl transferase subunit alpha
					CR921_RS02070	S-ribosylhomocysteine lyase
					CR921_RS02075	type I restriction endonuclease subunit S
					CR921_RS02080	restriction endonuclease
					CR921_RS02085	pseudo
					CR921_RS02090	glutamate decarboxylase
					CR921_RS02095	amino acid permease
					CR921_RS02100	hypothetical protein
					CR921_RS02105	IS6 family transposase
					CR921_RS02110	cytoplasmic protein
					CR921_RS02115	hypothetical protein
					CR921_RS02120	hypothetical protein
					CR921_RS02125	AbrB/MazE/SpoVT family DNA-binding domain-containing protein
					CR921_RS02130	pseudo
3	14,916	509,953	524,869	45.2	CR921_RS02755	hypothetical protein

					CR921_RS02760	hypothetical protein
					CR921_RS02765	hypothetical protein
					CR921_RS02770	hypothetical protein
					CR921_RS02775	hypothetical protein
					CR921_RS02780	hypothetical protein
					CR921_RS02785	hypothetical protein
					CR921_RS02790	hypothetical protein
					CR921_RS02795	hypothetical protein
					ltrA	group II intron reverse transcriptase/maturase
					CR921_RS02805	pseudo
					CR921_RS02810	hypothetical protein
					CR921_RS02815	hypothetical protein
					CR921_RS02820	plasmid mobilization relaxosome protein MobC
					CR921_RS02825	relaxase
4	5,812	807,545	813,357	39.5	CR921_RS04255	hypothetical protein
					CR921_RS04260	cysteine synthase family protein
					CR921_RS04265	cystathionine gamma-synthase
					CR921_RS04270	serine acetyltransferase
					CR921_RS04275	XRE family transcriptional regulator
					CR921_RS04280	pseudo
					CR921_RS04285	pseudo
					CR921_RS04290	DNA methyltransferase
5	5,144	860,300	865,444	29.4	CR921_RS04535	MutR family transcriptional regulator
					CR921_RS04540	pseudo
					CR921_RS04545	radical SAM protein
					CR921_RS04550	agmatinase
					CR921_RS04555	MFS transporter
6	4,668	895,089	899,757	33.6	CR921_RS04730	bacteriocin immunity protein
					CR921_RS04735	bacteriocin BlpK
					CR921_RS04740	hypothetical protein
					CR921_RS04745	ComC/BlpC family peptide pheromone/bacteriocin
					CR921_RS04750	bacteriocin immunity protein
					CR921_RS04755	hypothetical protein
					CR921_RS04760	hypothetical protein
					CR921_RS04765	ComC/BlpC family peptide pheromone/bacteriocin
					CR921_RS04770	pseudo
					CR921_RS04775	bacteriocin leader domain-containing protein
					CR921_RS04780	hypothetical protein
					CR921_RS04785	class IIb bacteriocin, lactobin A/cerein 7B family
					CR921_RS04790	hypothetical protein
7	3,555	1,012,855	1,016,410	37.4	CR921_RS05390	IS200/IS605 family transposase
					CR921_RS05395	50S ribosomal protein L20
					CR921_RS05400	50S ribosomal protein L35
					CR921_RS05405	translation initiation factor IF-3
					CR921_RS05410	(d)CMP kinase
					CR921_RS05415	LysM domain-containing protein
					CR921_RS05420	ferredoxin
					CR921_RS05425	hypothetical protein
8	4,426	1,048,971	1,053,397	39.0	CR921_RS05595	DUF3492 domain-containing protein
					CR921_RS05600	DUF2194 domain-containing protein
					CR921_RS05605	hypothetical protein

					CR921_RS05610	hypothetical protein
9	5,607	1,209,343	1,214,950	26.4	CR921_RS06410	MFS transporter permease
					CR921_RS06415	oligoendopeptidase F
					CR921_RS06420	helix-turn-helix domain-containing protein
					CR921_RS06425	NUDIX hydrolase
					CR921_RS06430	hypothetical protein
10	5,038	1,233,001	1,238,039	40.2	CR921_RS06530	pseudo
					CR921_RS06535	pseudo
					CR921_RS06540	pseudo
					CR921_RS06545	amino acid ABC transporter ATP-binding protein
11	9,369	1,496,917	1,506,286	34.4	CR921_RS07850	restriction endonuclease
					CR921_RS07855	SIR2 family protein
					CR921_RS07860	restriction endonuclease
					CR921_RS07865	N-6 DNA methylase
12	39,200	1,754,050	1,793,250	35.6	CR921_RS09150	hypothetical protein
					CR921_RS09155	hypothetical protein
					CR921_RS09160	helix-turn-helix domain-containing protein
					CR921_RS09165	IS30 family transposase
					CR921_RS09170	50S ribosomal protein L17
					CR921_RS09175	DNA-directed RNA polymerase subunit alpha
					CR921_RS09180	30S ribosomal protein S11
					CR921_RS09185	30S ribosomal protein S13
					CR921_RS09190	50S ribosomal protein L36
					CR921_RS09195	translation initiation factor IF-1
					CR921_RS09200	adenylate kinase
					CR921_RS09205	preprotein translocase subunit SecY
					CR921_RS09210	50S ribosomal protein L15
					CR921_RS09215	50S ribosomal protein L30
					CR921_RS09220	30S ribosomal protein S5
					CR921_RS09225	50S ribosomal protein L18
					CR921_RS09230	50S ribosomal protein L6
					CR921_RS09235	30S ribosomal protein S8
					CR921_RS09240	30S ribosomal protein S14 type Z
					CR921_RS09245	50S ribosomal protein L5
					CR921_RS09250	50S ribosomal protein L24
					CR921_RS09255	50S ribosomal protein L14
					CR921_RS09260	30S ribosomal protein S17
					CR921_RS09265	50S ribosomal protein L29
					CR921_RS09270	50S ribosomal protein L16
					CR921_RS09275	30S ribosomal protein S3
					CR921_RS09280	50S ribosomal protein L22
					CR921_RS09285	30S ribosomal protein S19
					CR921_RS09290	50S ribosomal protein L2
					CR921_RS09295	50S ribosomal protein L23
					CR921_RS09300	50S ribosomal protein L4
					CR921_RS09305	50S ribosomal protein L3
					CR921_RS09310	30S ribosomal protein S10
					CR921_RS09315	Holliday junction branch migration DNA helicase RuvB
					CR921_RS09320	acetyltransferase
					CR921_RS09325	membrane protein
					CR921_RS09330	low molecular weight phosphotyrosine protein phosphatase

					CR921_RS09335	adenylosuccinate synthetase
					CR921_RS09340	pseudo
					CR921_RS09345	MFS transporter
					CR921_RS09350	bacteriocin biosynthesis protein
					CR921_RS09355	pseudo
					CR921_RS09360	helix-turn-helix domain-containing protein
					CR921_RS09365	hypothetical protein
					CR921_RS09370	ABC transporter ATP-binding protein
					CR921_RS09375	MutR family transcriptional regulator
					CR921_RS09380	radical SAM protein
					CR921_RS09385	ABC transporter ATP-binding protein
					CR921_RS09390	XRE family transcriptional regulator
					CR921_RS09395	pseudo
					CR921_RS09400	DUF3173 domain-containing protein

Strain JIM8232						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	5,188	71,523	76,711	32.5	STH8232_RS00470	site-specific integrase
					STH8232_RS00475	DUF3173 domain-containing protein
					STH8232_RS00480	hypothetical protein
					STH8232_RS00485	replication initiation protein
					STH8232_RS00490	hypothetical protein
					STH8232_RS00495	XRE family transcriptional regulator
					STH8232_RS00500	glutamine amidotransferase
2	8,550	114,701	123,251	30.4	STH8232_RS00685	aminotransferase class V-fold PLP-dependent enzyme
					STH8232_RS00690	KR domain-containing protein
					STH8232_RS00695	hypothetical protein
					STH8232_RS00700	ABC transporter ATP-binding protein
					STH8232_RS00705	ABC transporter permease
					STH8232_RS00710	TetR/AcrR family transcriptional regulator
3	13,898	125,465	139,363	33.9	STH8232_RS00725	site-specific integrase
					STH8232_RS00730	DNA-binding protein
					STH8232_RS00735	plasmid replication protein
					STH8232_RS00740	hypothetical protein
					STH8232_RS00745	cell division protein FtsK
					STH8232_RS00750	hypothetical protein
					STH8232_RS00755	hypothetical protein
					STH8232_RS00760	MFS transporter
					STH8232_RS00765	NAD(P)-dependent alcohol dehydrogenase
					STH8232_RS00770	Rrf2 family transcriptional regulator
					STH8232_RS00775	HAD family hydrolase
					STH8232_RS00780	carboxymuconolactone decarboxylase family protein
					STH8232_RS00785	histidine phosphatase family protein
					STH8232_RS00790	hypothetical protein
					STH8232_RS00795	XRE family transcriptional regulator
					STH8232_RS00800	pseudo
					STH8232_RS00805	hypothetical protein
4	14,308	566,778	581,086	31.1	STH8232_RS03045	XRE family transcriptional regulator
					STH8232_RS03050	flavoprotein

					STH8232_RS03055	pyridoxal-dependent decarboxylase
					STH8232_RS03060	hypothetical protein
					STH8232_RS03065	MFS transporter
					STH8232_RS03070	IS6 family transposase
					STH8232_RS03075	XRE family transcriptional regulator
					STH8232_RS03080	DUF3173 domain-containing protein
					STH8232_RS03085	site-specific integrase
					STH8232_RS03090	ISL3 family transposase
					STH8232_RS10340	hypothetical protein
					STH8232_RS03095	ABC transporter ATP-binding protein
					STH8232_RS03100	peptide ABC transporter permease
5	8,753	906,541	915,294	32.4	STH8232_RS04730	hypothetical protein
					STH8232_RS04735	recombinase
					STH8232_RS04740	hypothetical protein
					STH8232_RS04745	zinc ribbon domain-containing protein
					STH8232_RS04750	pseudo
					STH8232_RS04755	DUF1643 domain-containing protein
					STH8232_RS04760	hypothetical protein
					STH8232_RS04765	DUF927 domain-containing protein
					STH8232_RS04770	DUF3173 domain-containing protein
					STH8232_RS04775	site-specific integrase
					STH8232_RS04780	glutamine-hydrolyzing GMP synthase
6	5,143	937,813	942,956	29.3	STH8232_RS04885	pseudo
					STH8232_RS04890	Fe-S oxidoreductase
					STH8232_RS04895	radical SAM protein
					STH8232_RS04900	agmatinase
					STH8232_RS04905	MFS transporter
7	8,390	1,193,475	1,201,865	41.6	STH8232_RS06205	phosphoribosyl-ATP pyrophosphatase
					STH8232_RS06210	phosphoribosyl-AMP cyclohydrolase
					STH8232_RS06215	imidazole glycerol phosphate synthase subunit HisF
					STH8232_RS06220	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase
					STH8232_RS06225	imidazole glycerol phosphate synthase subunit HisH
					hisB	imidazoleglycerol-phosphate dehydratase
					STH8232_RS06235	histidinol dehydrogenase
					STH8232_RS06240	ATP phosphoribosyltransferase
					STH8232_RS06245	ATP phosphoribosyltransferase regulatory subunit
					STH8232_RS06250	histidinol-phosphate aminotransferase
					STH8232_RS06255	hypothetical protein
8	5,607	1,285,617	1,291,224	26.5	STH8232_RS06675	MFS transporter permease
					STH8232_RS06680	oligoendopeptidase F
					STH8232_RS06685	helix-turn-helix domain-containing protein
					STH8232_RS06690	NUDIX hydrolase
					STH8232_RS06695	hypothetical protein
9	8,118	1,333,414	1,341,532	31.9	STH8232_RS06935	sodium transporter
					STH8232_RS06940	pseudo
					STH8232_RS06945	pseudo
					STH8232_RS06950	transporter
					STH8232_RS06955	KxxxW cyclic peptide radical SAM maturase
					STH8232_RS10725	KxxxW-cyclized peptide pheromone

					STH8232_RS06960	helix-turn-helix domain-containing protein
					STH8232_RS10730	pseudo
					STH8232_RS06975	hypothetical protein
10	6,193	1,494,953	1,501,146	33.3	STH8232_RS07725	helix-turn-helix domain-containing protein
					STH8232_RS07730	radical SAM protein
					STH8232_RS07735	radical SAM/SPASM domain-containing protein
					STH8232_RS07740	hypothetical protein
					STH8232_RS07745	ABC transporter ATP-binding protein
					STH8232_RS07750	hypothetical protein
					STH8232_RS07755	hypothetical protein
11	4,477	1,606,273	1,610,750	36.8	STH8232_RS08305	hypothetical protein
					STH8232_RS08310	hypothetical protein
					STH8232_RS08315	ComC/BlpC family peptide pheromone/bacteriocin
					STH8232_RS08320	hypothetical protein
					STH8232_RS10845	hypothetical protein
					STH8232_RS08330	ComC/BlpC family peptide pheromone/bacteriocin
					STH8232_RS10850	hypothetical protein
					STH8232_RS08335	hypothetical protein
					STH8232_RS08340	ComC/BlpC family peptide pheromone/bacteriocin
					STH8232_RS08345	DNA-binding response regulator
					STH8232_RS08350	GHKL domain-containing protein
12	58,208	1,808,538	1,866,746	35.3	STH8232_RS09370	hypothetical protein
					STH8232_RS09375	helix-turn-helix domain-containing protein
					STH8232_RS09380	IS30 family transposase
					STH8232_RS09385	hypothetical protein
					STH8232_RS09390	hypothetical protein
					STH8232_RS09400	hypothetical protein
					STH8232_RS09405	IS5 family transposase ISSth7
					STH8232_RS09410	hypothetical protein
					STH8232_RS09415	toll/interleukin-1 receptor domain-containing protein
					STH8232_RS10920	restriction endonuclease subunit S
					STH8232_RS09425	type I restriction-modification system subunit M
					STH8232_RS09430	type I restriction endonuclease subunit R
					STH8232_RS10925	DNA-binding protein
					STH8232_RS09440	IS256 family transposase
					STH8232_RS10930	site-specific integrase
					STH8232_RS09450	fructose-1,6-bisphosphate aldolase, class II
					STH8232_RS09460	glutamate--cysteine ligase
					STH8232_RS09465	hypothetical protein
					STH8232_RS09470	hypothetical protein
					STH8232_RS10935	helix-turn-helix domain-containing protein
					STH8232_RS09475	IS30 family transposase
					STH8232_RS09480	50S ribosomal protein L17
					STH8232_RS09485	DNA-directed RNA polymerase subunit alpha
					STH8232_RS09490	30S ribosomal protein S11
					STH8232_RS09495	30S ribosomal protein S13
					STH8232_RS10940	50S ribosomal protein L36
					STH8232_RS09500	translation initiation factor IF-1
					STH8232_RS09505	adenylate kinase
					STH8232_RS09510	preprotein translocase subunit SecY
					STH8232_RS09515	50S ribosomal protein L15

					STH8232_RS09520	50S ribosomal protein L30
					STH8232_RS09525	30S ribosomal protein S5
					STH8232_RS09530	50S ribosomal protein L18
					STH8232_RS09535	50S ribosomal protein L6
					STH8232_RS09540	30S ribosomal protein S8
					rpsN	30S ribosomal protein S14 type Z
					STH8232_RS09550	50S ribosomal protein L5
					STH8232_RS09555	50S ribosomal protein L24
					STH8232_RS09560	50S ribosomal protein L14
					STH8232_RS09565	30S ribosomal protein S17
					STH8232_RS09570	50S ribosomal protein L29
					STH8232_RS09575	50S ribosomal protein L16
					STH8232_RS09580	30S ribosomal protein S3
					STH8232_RS09585	50S ribosomal protein L22
					STH8232_RS09590	30S ribosomal protein S19
					STH8232_RS09595	50S ribosomal protein L2
					STH8232_RS09600	50S ribosomal protein L23
					STH8232_RS09605	50S ribosomal protein L4
					STH8232_RS09610	50S ribosomal protein L3
					STH8232_RS09615	30S ribosomal protein S10
					STH8232_RS09620	Holliday junction branch migration DNA helicase RuvB
					STH8232_RS09625	acetyltransferase
					STH8232_RS09630	membrane protein
					STH8232_RS09635	low molecular weight phosphotyrosine protein phosphatase
					STH8232_RS09640	adenylosuccinate synthetase
					STH8232_RS09645	adenylate kinase
					STH8232_RS09650	MFS transporter
					STH8232_RS09655	bacteriocin biosynthesis protein
					STH8232_RS09665	IS256-like element IS1191 family transposase
					STH8232_RS09670	helix-turn-helix domain-containing protein
					STH8232_RS09675	pseudo
					STH8232_RS09680	ABC transporter ATP-binding protein
					STH8232_RS09685	MutR family transcriptional regulator
					STH8232_RS09690	radical SAM protein
					STH8232_RS09695	ABC transporter ATP-binding protein
					STH8232_RS09700	XRE family transcriptional regulator
					STH8232_RS10945	hypothetical protein
					STH8232_RS10950	replication initiator protein
					STH8232_RS09715	DUF3173 domain-containing protein

Strain KLDS 3.1003						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	14,484	8,767	23,251	38.0	BEN15_RS09990	pseudo
					BEN15_RS00040	transposase
					BEN15_RS00045	pseudo
					BEN15_RS09995	hypothetical protein
					BEN15_RS00050	ABC transporter ATP-binding protein
					BEN15_RS10000	pseudo
					BEN15_RS00055	pseudo

					BEN15_RS00060	bifunctional glutamate--cysteine ligase GshA/glutathione synthetase GshB
					BEN15_RS10005	pseudo
					BEN15_RS00080	hypothetical protein
					BEN15_RS00085	DUF3114 domain-containing protein
					BEN15_RS00090	pseudo
					BEN15_RS00095	ABC transporter ATP-binding protein
					BEN15_RS00100	pseudo
					BEN15_RS00105	helix-turn-helix domain-containing protein
2	6,783	204,864	211,647	29.5	BEN15_RS10065	hypothetical protein
					BEN15_RS01005	FRG domain-containing protein
					BEN15_RS01010	hypothetical protein
					BEN15_RS01015	hypothetical protein
					BEN15_RS01020	pseudo
					BEN15_RS01025	dicarboxylate/amino acid:cation symporter
					BEN15_RS01030	alpha-L-glutamate ligase
					BEN15_RS01035	pseudo
3	11,428	371,688	383,116	38.1	BEN15_RS01820	aromatic acid exporter family protein
					BEN15_RS10105	pseudo
					BEN15_RS01830	phosphoglycerate kinase
					BEN15_RS01835	pseudo
					BEN15_RS01840	IS3 family transposase
					BEN15_RS01845	hypothetical protein
					BEN15_RS01850	transposase
					BEN15_RS10110	pseudo
					BEN15_RS01865	type I glyceraldehyde-3-phosphate dehydrogenase
					BEN15_RS01870	elongation factor G
					BEN15_RS01875	30S ribosomal protein S7
					BEN15_RS01880	30S ribosomal protein S12
					BEN15_RS01885	pur operon repressor
					BEN15_RS01890	HD domain-containing protein
					BEN15_RS01895	DNA recombination protein RmuC
4	10,253	499,893	510,146	28.8	BEN15_RS02605	DNA topology modulation protein FlaR
					BEN15_RS02610	pseudo
					BEN15_RS02615	helix-turn-helix domain-containing protein
					BEN15_RS02620	hypothetical protein
					BEN15_RS02625	ABC transporter ATP-binding protein
					BEN15_RS02630	MutR family transcriptional regulator
					BEN15_RS02635	radical SAM protein
					BEN15_RS02640	ABC transporter ATP-binding protein
5	19,588	1,209,205	1,228,793	38.7	BEN15_RS06465	conjugal transfer protein TraG
					BEN15_RS06470	hypothetical protein
					BEN15_RS06475	conjugal transfer protein TrbL
					BEN15_RS06480	PrgI family protein
					BEN15_RS06485	AAA family ATPase
					BEN15_RS06490	CHAP domain-containing protein
					BEN15_RS06495	pseudo
					BEN15_RS06500	YydF family exported signaling peptide
					BEN15_RS06505	IS30 family transposase
					BEN15_RS06510	peptide ABC transporter ATP-binding protein
					BEN15_RS06515	ABC transporter permease

					BEN15_RS06520	peptidylprolyl isomerase
					BEN15_RS06525	DUF3991 domain-containing protein
					BEN15_RS06530	hypothetical protein
					BEN15_RS06535	hypothetical protein
					BEN15_RS06540	hypothetical protein
					BEN15_RS06545	hypothetical protein
					BEN15_RS06550	plasmid mobilization relaxosome protein MobC
					BEN15_RS06555	relaxase
					BEN15_RS10450	hypothetical protein
					BEN15_RS06560	LytTR family transcriptional regulator
					BEN15_RS06565	DNA-binding response regulator
6	22,858	1,407,220	1,430,078	36.7	BEN15_RS07450	sensory protein
					BEN15_RS07455	pseudo
					BEN15_RS07460	pseudo
					BEN15_RS07465	pseudo
					BEN15_RS07470	IS6 family transposase
					BEN15_RS07480	hypothetical protein
					BEN15_RS07485	cold-shock protein
					BEN15_RS07490	cold-shock protein
					BEN15_RS07495	copper-translocating P-type ATPase
					BEN15_RS07500	pseudo
					BEN15_RS07505	pseudo
					BEN15_RS10595	hypothetical protein
					BEN15_RS07510	cysteine synthase family protein
					BEN15_RS07515	cystathionine gamma-synthase
					BEN15_RS07520	serine acetyltransferase
					BEN15_RS10600	pseudo
					BEN15_RS07525	pseudo
					BEN15_RS10605	pseudo
					BEN15_RS10610	pseudo
					BEN15_RS07535	pseudo
					BEN15_RS10615	pseudo
					BEN15_RS07540	sugar transporter
					BEN15_RS10620	pseudo
					BEN15_RS07545	type I restriction-modification system subunit R
					BEN15_RS07550	pseudo
					BEN15_RS10625	pseudo
					BEN15_RS07560	zinc ABC transporter substrate-binding protein AdcA
					BEN15_RS07565	hypothetical protein
7	19,054	1,482,261	1,501,315	38.2	BEN15_RS07845	DUF814 domain-containing protein
					BEN15_RS07850	site-specific integrase
					BEN15_RS07855	XRE family transcriptional regulator
					BEN15_RS07860	helix-turn-helix domain-containing protein
					BEN15_RS07865	pseudo
					BEN15_RS07870	hypothetical protein
					BEN15_RS07875	hypothetical protein
					BEN15_RS07880	DNA-binding protein
					BEN15_RS07885	hypothetical protein
					BEN15_RS07890	hypothetical protein
					BEN15_RS07895	virulence-associated protein E
					BEN15_RS07900	hypothetical protein

					BEN15_RS10685	pseudo
					BEN15_RS07905	pseudo
					BEN15_RS07910	hypothetical protein
					BEN15_RS07915	amino acid permease
					BEN15_RS07920	histidine decarboxylase, pyruvoyl type
					BEN15_RS07925	hypothetical protein
					BEN15_RS07930	type I restriction endonuclease subunit R
					BEN15_RS07935	type I restriction-modification system subunit M
					BEN15_RS07940	restriction endonuclease subunit S
8	7,441	1,576,539	1,583,980	27.4	BEN15_RS08320	hypothetical protein
					BEN15_RS08325	hypothetical protein
					BEN15_RS08330	hypothetical protein
					BEN15_RS08335	UDP-N-acetylglucosamine enolpyruvyl transferase
					BEN15_RS08340	pseudo
					BEN15_RS08345	tyrosine--tRNA ligase
9	5,344	1,614,385	1,619,729	30.3	BEN15_RS08515	glycosyltransferase
					BEN15_RS08520	pseudo
					BEN15_RS08525	pseudo
					BEN15_RS08530	glycosyltransferase family 2 protein
					BEN15_RS08535	oligosaccharide repeat unit polymerase
					BEN15_RS08540	DUF4422 domain-containing protein
10	5,607	1,832,431	1,838,038	26.5	BEN15_RS09660	transporter
					BEN15_RS09665	oligoendopeptidase F
					BEN15_RS09670	helix-turn-helix domain-containing protein
					BEN15_RS09675	NUDIX hydrolase
					BEN15_RS09680	hypothetical protein
11	8,258	1,850,992	1,859,250	35.7	BEN15_RS09750	amino acid ABC transporter substrate-binding protein
					BEN15_RS09755	adenylosuccinate lyase
					BEN15_RS09760	amino acid ABC transporter ATP-binding protein
					BEN15_RS09765	amino acid ABC transporter permease
					BEN15_RS09770	N-acetyltransferase
					BEN15_RS09775	hypothetical protein
					BEN15_RS09780	pseudo
					BEN15_RS10825	pseudo
					BEN15_RS09800	amidohydrolase
12	4,378	1,886,857	1,891,235	30.6	BEN15_RS09955	pseudo
					BEN15_RS09960	MFS transporter
					BEN15_RS09965	KxxxW cyclic peptide radical SAM maturase
					BEN15_RS10865	KxxxW-cyclized peptide pheromone
					BEN15_RS09970	helix-turn-helix domain-containing protein

KLDS SM						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	7,212	54,248	61,460	37.0	A9497_RS00275	hypothetical protein
					A9497_RS00280	hypothetical protein
					A9497_RS00285	DNA primase
					A9497_RS00290	hypothetical protein
					A9497_RS00295	hypothetical protein
					A9497_RS09775	pseudo

					A9497_RS00300	hypothetical protein
					A9497_RS00305	hypothetical protein
					A9497_RS00310	hypothetical protein
					A9497_RS00315	XRE family transcriptional regulator
					A9497_RS09780	XRE family transcriptional regulator
					A9497_RS00325	site-specific integrase
					A9497_RS00330	hypothetical protein
2	6,466	169,624	176,090	31.1	A9497_RS00875	MutR family transcriptional regulator
					A9497_RS00880	Fe-S oxidoreductase
					A9497_RS00885	IS256 family transposase
					A9497_RS00890	radical SAM protein
					A9497_RS00895	agmatinase
					A9497_RS00900	MFS transporter
3	8,409	271,805	280,214	34.7	A9497_RS01390	type I-E CRISPR-associated endonuclease Cas1
					A9497_RS01395	type I-E CRISPR-associated protein Cas6/Cse3/CasE
					A9497_RS01400	type I-E CRISPR-associated protein Cas5/CasD
					A9497_RS01405	type I-E CRISPR-associated protein Cas7/Cse4/CasC
					A9497_RS01410	type I-E CRISPR-associated protein Cse2/CasB
					A9497_RS01415	type I-E CRISPR-associated protein Cse1/CasA
					A9497_RS01420	CRISPR-associated helicase/endonuclease Cas3
					A9497_RS01425	DUF2207 domain-containing protein
4	5,607	516,264	521,871	26.4	A9497_RS02645	MFS transporter permease
					A9497_RS02650	oligoendopeptidase F
					A9497_RS02655	helix-turn-helix domain-containing protein
					A9497_RS02660	NUDIX hydrolase
					A9497_RS02665	hypothetical protein
5	7,681	562,306	569,987	30.7	A9497_RS02885	ABC transporter ATP-binding protein
					A9497_RS02890	hypothetical protein
					A9497_RS02895	hypothetical protein
					A9497_RS02900	sodium transporter
					A9497_RS02905	pseudo
					A9497_RS02910	SPASM domain-containing protein
					A9497_RS02915	transporter
					A9497_RS02920	pseudo
					A9497_RS10075	KxxxW-cyclized peptide pheromone
					A9497_RS02925	helix-turn-helix domain-containing protein
6	4,070	1,050,630	1,054,700	38.4	A9497_RS05370	type I restriction-modification system subunit M
					A9497_RS05375	pseudo
					A9497_RS05380	DNA-binding protein
7	38,269	1,060,334	1,098,603	35.6	A9497_RS10250	helix-turn-helix domain-containing protein
					A9497_RS05420	IS30 family transposase
					A9497_RS05425	50S ribosomal protein L17
					A9497_RS05430	DNA-directed RNA polymerase subunit alpha
					A9497_RS05435	30S ribosomal protein S11
					A9497_RS05440	30S ribosomal protein S13
					A9497_RS10255	50S ribosomal protein L36
					A9497_RS05445	translation initiation factor IF-1
					A9497_RS05450	adenylate kinase
					A9497_RS05455	preprotein translocase subunit SecY
					A9497_RS05460	50S ribosomal protein L15
					A9497_RS05465	50S ribosomal protein L30

					A9497_RS05470	30S ribosomal protein S5
					A9497_RS05475	50S ribosomal protein L18
					A9497_RS05480	50S ribosomal protein L6
					A9497_RS05485	30S ribosomal protein S8
					rpsN	30S ribosomal protein S14 type Z
					A9497_RS05495	50S ribosomal protein L5
					A9497_RS05500	50S ribosomal protein L24
					A9497_RS05505	50S ribosomal protein L14
					A9497_RS05510	30S ribosomal protein S17
					A9497_RS05515	50S ribosomal protein L29
					A9497_RS05520	50S ribosomal protein L16
					A9497_RS05525	30S ribosomal protein S3
					A9497_RS05530	50S ribosomal protein L22
					A9497_RS05535	30S ribosomal protein S19
					A9497_RS05540	50S ribosomal protein L2
					A9497_RS05545	50S ribosomal protein L23
					A9497_RS05550	50S ribosomal protein L4
					A9497_RS05555	50S ribosomal protein L3
					A9497_RS05560	30S ribosomal protein S10
					A9497_RS05565	Holliday junction branch migration DNA helicase RuvB
					A9497_RS05570	acyltransferase
					A9497_RS05575	membrane protein
					A9497_RS05580	low molecular weight phosphotyrosine protein phosphatase
					A9497_RS05585	adenylosuccinate synthetase
					A9497_RS05590	pseudo
					A9497_RS05595	MFS transporter
					A9497_RS05600	bacteriocin biosynthesis protein
					A9497_RS05610	pseudo
					A9497_RS05615	helix-turn-helix domain-containing protein
					A9497_RS05620	hypothetical protein
					A9497_RS05625	ABC transporter ATP-binding protein
					A9497_RS05630	MutR family transcriptional regulator
					A9497_RS05635	radical SAM protein
					A9497_RS05640	ABC transporter ATP-binding protein
					A9497_RS05645	pseudo
					A9497_RS10260	hypothetical protein
					A9497_RS10265	replication initiator protein
					A9497_RS05660	DUF3173 domain-containing protein
8	10,748	1,119,811	1,130,559	34.5	A9497_RS05795	alkaline shock response membrane anchor protein AmaP
					A9497_RS05800	DUF2273 domain-containing protein
					A9497_RS05805	Asp23/Gls24 family envelope stress response protein
					A9497_RS05810	CsbD family protein
					A9497_RS05815	phosphohydrolase
					A9497_RS05820	IS256 family transposase
					A9497_RS10280	PadR family transcriptional regulator
					A9497_RS10285	PadR family transcriptional regulator
					A9497_RS05830	hypothetical protein
					A9497_RS05835	hypothetical protein
					A9497_RS05840	YhgE/Pip domain-containing protein
					A9497_RS05845	TetR/AcrR family transcriptional regulator
					A9497_RS05850	DUF1304 domain-containing protein

					A9497_RS05855	30S ribosomal protein S4
9	25,787	1,179,539	1,205,326	42.1	A9497_RS06090	transposase
					A9497_RS06095	IS3 family transposase
					A9497_RS06205	rod shape-determining protein MreC
					A9497_RS06210	rod shape-determining protein MreD
					A9497_RS06215	CHAP domain-containing protein
					A9497_RS06220	ribose-phosphate diphosphokinase
					A9497_RS10295	pseudo
					A9497_RS06230	pyridoxal phosphate-dependent aminotransferase
					A9497_RS06235	DNA repair protein RecO
					A9497_RS06240	phosphate acyltransferase PlsX
					A9497_RS06245	acyl carrier protein
					A9497_RS06250	phosphoribosylaminoimidazolesuccinocarboxamide synthase
					A9497_RS06255	phosphoribosylformylglycinamidine synthase
					A9497_RS06260	amidophosphoribosyltransferase
					A9497_RS06265	phosphoribosylformylglycinamidine cyclo-ligase
					A9497_RS06270	phosphoribosylglycinamide formyltransferase
					A9497_RS06275	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/inosine monophosphate cyclohy
10	13,289	1,262,534	1,275,823	32.9	A9497_RS06620	hypothetical protein
					A9497_RS06625	2-dehydropantoate 2-reductase
					A9497_RS06630	50S ribosomal protein L13
					A9497_RS06635	30S ribosomal protein S9
					A9497_RS06640	ISL3 family transposase
					A9497_RS10325	pseudo
					A9497_RS06645	helix-turn-helix transcriptional regulator
					A9497_RS06650	hypothetical protein
					A9497_RS06655	pseudo
					A9497_RS06660	lantibiotic biosynthesis protein
					A9497_RS06665	lantibiotic transporter
					A9497_RS06670	pseudo
					A9497_RS06675	pseudo
					A9497_RS06680	DNA-binding protein
11	10,373	1,659,779	1,670,152	36.8	A9497_RS08740	ABC transporter substrate-binding protein
					A9497_RS10480	hypothetical protein
					A9497_RS10485	pseudo
					A9497_RS08750	macrolide ABC transporter
					A9497_RS08755	ABC transporter permease, truncated
					A9497_RS08760	50S ribosomal protein L10
					A9497_RS08765	50S ribosomal protein L7/L12
					A9497_RS08775	pseudo
					A9497_RS10490	hypothetical protein
					A9497_RS08780	pseudo
					A9497_RS08785	ABC transporter permease
					A9497_RS10495	pseudo
					A9497_RS08800	ABC transporter permease
					A9497_RS10500	hypothetical protein
					A9497_RS08805	hypothetical protein
					A9497_RS08810	hypothetical protein
					A9497_RS08815	hypothetical protein
12	4,410	1,791,194	1,795,604	36.8	A9497_RS09420	transcriptional regulator

					A9497_RS09425	pseudo
					A9497_RS09430	hypothetical protein
					A9497_RS09435	hypothetical protein
					A9497_RS09440	hypothetical protein
					A9497_RS09445	pseudo
					A9497_RS10550	hypothetical protein
					A9497_RS09450	IS3 family transposase
					A9497_RS10555	pseudo
13	5,579	1,845,638	1,851,217	29.2	A9497_RS09705	hypothetical protein
					A9497_RS09710	hypothetical protein
					A9497_RS09715	hypothetical protein
					A9497_RS09720	restriction endonuclease subunit S
					A9497_RS09725	NTPase

Strain LMD-9						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	9,373	747,073	756,446	36.7	STER_RS04000	hypothetical protein
					STER_RS04005	hypothetical protein
					STER_RS04010	hypothetical protein
					STER_RS04015	hypothetical protein
					STER_RS04020	hypothetical protein
					STER_RS04025	DNA primase
					STER_RS04030	hypothetical protein
					STER_RS04035	hypothetical protein
					STER_RS04040	hypothetical protein
					STER_RS04045	hypothetical protein
					STER_RS04050	hypothetical protein
					STER_RS04055	pseudo
					STER_RS04060	XRE family transcriptional regulator
					STER_RS04065	XRE family transcriptional regulator
					STER_RS04070	site-specific integrase
2	17,203	812,146	829,349	37.7	STER_RS04365	copper-translocating P-type ATPase
					STER_RS04370	IS256 family transposase
					STER_RS04375	pseudo
					STER_RS10300	hypothetical protein
					STER_RS04380	cysteine synthase family protein
					STER_RS04385	cystathionine gamma-synthase
					STER_RS04390	serine acetyltransferase
					STER_RS10305	pseudo
					STER_RS04395	IS256 family transposase
					STER_RS10310	pseudo
					STER_RS04410	DNA methyltransferase
					STER_RS10315	pseudo
					STER_RS04415	sugar transporter
					STER_RS10320	pseudo
					STER_RS04420	hypothetical protein
					STER_RS04425	pseudo
					STER_RS04430	PrsW family intramembrane metalloprotease
					STER_RS04435	zinc ABC transporter substrate-binding protein AdcA

					STER_RS04440	hypothetical protein
3	5,144	865,021	870,165	29.3	STER_RS04620	MutR family transcriptional regulator
					STER_RS04625	Fe-S oxidoreductase
					STER_RS04630	radical SAM protein
					STER_RS04635	agmatinase
					STER_RS04640	MFS transporter
4	7,922	981,992	989,914	32.3	STER_RS05255	pseudo
					STER_RS05260	cell wall biosynthesis glycosyltransferase
					STER_RS05265	polysaccharide polymerase
					STER_RS05270	glycosyl transferase family 1
					STER_RS05275	cell wall biosynthesis glycosyltransferase
					STER_RS05280	glycosyltransferase family 1 protein
					STER_RS05285	multidrug MFS transporter
					STER_RS05290	UDP-N-acetylglucosamine--LPS N-acetylglucosamine transferase
					STER_RS10440	pseudo
5	4,426	1,043,027	1,047,453	38.9	STER_RS05585	DUF3492 domain-containing protein
					STER_RS05590	DUF2194 domain-containing protein
					STER_RS05595	hypothetical protein
					STER_RS10460	membrane protein
6	5,607	1,203,277	1,208,884	26.4	STER_RS06395	MFS transporter permease
					STER_RS06400	oligoendopeptidase F
					STER_RS06405	helix-turn-helix domain-containing protein
					STER_RS06410	NUDIX hydrolase
					STER_RS06415	hypothetical protein
7	9,259	1,254,918	1,264,177	31.5	STER_RS06655	ABC transporter ATP-binding protein
					STER_RS06660	hypothetical protein
					STER_RS06665	hypothetical protein
					STER_RS06670	sodium transporter
					STER_RS06675	pseudo
					STER_RS06680	SPASM domain-containing protein
					STER_RS06685	transporter
					STER_RS06690	KxxxW cyclic peptide radical SAM maturase
					STER_RS10575	KxxxW-cyclized peptide pheromone
					STER_RS06695	helix-turn-helix domain-containing protein
					STER_RS10580	pseudo
8	4,495	1,533,420	1,537,915	36.9	STER_RS08040	hypothetical protein
					STER_RS08045	hypothetical protein
					STER_RS08050	ComC/BlpC family peptide pheromone/bacteriocin
					STER_RS08055	hypothetical protein
					STER_RS10690	hypothetical protein
					STER_RS08065	ComC/BlpC family peptide pheromone/bacteriocin
					STER_RS10695	hypothetical protein
					STER_RS08070	hypothetical protein
					STER_RS08075	ComC/BlpC family peptide pheromone/bacteriocin
					STER_RS08080	DNA-binding response regulator
					STER_RS08085	GHKL domain-containing protein
9	28,974	1,710,761	1,739,735	40.9	STER_RS08990	prepilin-type N-terminal cleavage/methylation domain-containing protein
					STER_RS08995	competence protein CglB
					STER_RS09000	competence protein CglA
					STER_RS09005	DUF1033 domain-containing protein

					STER_RS09010	DNA-directed RNA polymerase subunit beta'
					STER_RS09015	DNA-directed RNA polymerase subunit beta
					STER_RS09020	penicillin-binding protein
					STER_RS09025	tyrosine--tRNA ligase
					STER_RS09030	ketol-acid reductoisomerase
					STER_RS09035	acetolactate synthase small subunit
					STER_RS09040	acetolactate synthase, large subunit, biosynthetic type
					STER_RS09045	dihydroxy-acid dehydratase
					STER_RS09050	ISL3 family transposase
					STER_RS09055	DAK2 domain-containing protein
					STER_RS09060	Asp23/Gls24 family envelope stress response protein
					STER_RS09065	MATE family efflux transporter
					STER_RS09070	threonine synthase
					STER_RS09075	aldehyde dehydrogenase
					STER_RS09080	aldehyde dehydrogenase
					STER_RS09085	alcohol dehydrogenase
					STER_RS09090	alcohol dehydrogenase
					STER_RS09095	hypothetical protein
10	39,145	1,753,132	1,792,274	35.5	STER_RS09200	hypothetical protein
					STER_RS09205	hypothetical protein
					STER_RS10780	helix-turn-helix domain-containing protein
					STER_RS09210	IS30 family transposase
					STER_RS09215	50S ribosomal protein L17
					STER_RS09220	DNA-directed RNA polymerase subunit alpha
					STER_RS09225	30S ribosomal protein S11
					STER_RS09230	30S ribosomal protein S13
					STER_RS10785	50S ribosomal protein L36
					STER_RS09235	translation initiation factor IF-1
					STER_RS09240	adenylate kinase
					STER_RS09245	preprotein translocase subunit SecY
					STER_RS09250	50S ribosomal protein L15
					STER_RS09255	50S ribosomal protein L30
					STER_RS09260	30S ribosomal protein S5
					STER_RS09265	50S ribosomal protein L18
					STER_RS09270	50S ribosomal protein L6
					STER_RS09275	30S ribosomal protein S8
					rpsN	30S ribosomal protein S14 type Z
					STER_RS09285	50S ribosomal protein L5
					STER_RS09290	50S ribosomal protein L24
					STER_RS09295	50S ribosomal protein L14
					STER_RS09300	30S ribosomal protein S17
					STER_RS09305	50S ribosomal protein L29
					STER_RS09310	50S ribosomal protein L16
					STER_RS09315	30S ribosomal protein S3
					STER_RS09320	50S ribosomal protein L22
					STER_RS09325	30S ribosomal protein S19
					STER_RS09330	50S ribosomal protein L2
					STER_RS09335	50S ribosomal protein L23
					STER_RS09340	50S ribosomal protein L4
					STER_RS09345	50S ribosomal protein L3
					STER_RS09350	30S ribosomal protein S10

					STER_RS09355	Holliday junction branch migration DNA helicase RuvB
					STER_RS09360	acetyltransferase
					STER_RS09365	membrane protein
					STER_RS09370	low molecular weight phosphotyrosine protein phosphatase
					STER_RS09375	adenylosuccinate synthetase
					STER_RS09380	pseudo
					STER_RS09385	MFS transporter
					STER_RS09390	bacteriocin biosynthesis protein
					STER_RS09400	pseudo
					STER_RS09405	helix-turn-helix domain-containing protein
					STER_RS09410	hypothetical protein
					STER_RS09415	ABC transporter ATP-binding protein
					STER_RS09420	MutR family transcriptional regulator
					STER_RS09425	radical SAM protein
					STER_RS09430	ABC transporter ATP-binding protein
					STER_RS09435	XRE family transcriptional regulator
					STER_RS10790	hypothetical protein
					STER_RS10795	hypothetical protein
					STER_RS09450	DUF3173 domain-containing protein

Strain LMG 18311						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	13,260	100,771	114,031	31.8	STU_RS19020	transcriptional regulator
					STU_RS10080	pseudo
					STU_RS10085	2-dehydropantoate 2-reductase
					STU_RS10090	50S ribosomal protein L13
					STU_RS10095	30S ribosomal protein S9
					STU_RS19025	hypothetical protein
					STU_RS19030	pseudo
					STU_RS10100	helix-turn-helix transcriptional regulator
					STU_RS10105	hypothetical protein
					STU_RS10110	lantibiotic biosynthesis protein
					STU_RS10115	lantibiotic biosynthesis protein
					STU_RS10120	pseudo
					STU_RS10125	pseudo
					STU_RS10130	integrase
					STU_RS10135	Fis family transcriptional regulator
					STU_RS10140	pseudo
2	5,652	664,335	669,987	30.3	STU_RS13015	type I restriction endonuclease subunit R
					STU_RS13020	hypothetical protein
					STU_RS13025	hypothetical protein
					STU_RS13030	restriction endonuclease subunit S
					STU_RS13035	NTPase
3	15,660	770,113	785,773	37.7	STU_RS13555	cold-shock protein
					STU_RS13560	cold-shock protein
					STU_RS13565	IS256 family transposase
					STU_RS13570	copper-translocating P-type ATPase
					STU_RS13575	pseudo
					STU_RS13580	type I restriction endonuclease subunit R

					STU_RS19315	hypothetical protein
					STU_RS13585	cysteine synthase family protein
					STU_RS13590	cystathionine gamma-synthase
					STU_RS13595	serine acetyltransferase
					STU_RS19320	pseudo
					STU_RS13600	IS256 family transposase
					STU_RS19325	pseudo
					STU_RS19330	pseudo
					STU_RS13615	type I restriction-modification system methyltransferase subunit, truncated
					STU_RS19335	pseudo
					STU_RS13620	sugar transporter
					STU_RS19340	pseudo
					STU_RS13625	type I restriction-modification system subunit R
4	6,263	833,745	840,008	30.1	STU_RS13870	hypothetical protein
					STU_RS13875	XRE family transcriptional regulator
					STU_RS13880	MutR family transcriptional regulator
					STU_RS13885	Fe-S oxidoreductase
					STU_RS13890	radical SAM protein
					STU_RS13895	agmatinase
					STU_RS13900	MFS transporter
5	10,849	965,123	975,972	30.1	STU_RS14555	hypothetical protein
					STU_RS14560	exopolysaccharide biosynthesis protein
					STU_RS14565	exopolysaccharide biosynthesis protein
					STU_RS14570	alpha-1,2-fucosyltransferase
					STU_RS14575	exopolysaccharide biosynthesis protein
					STU_RS18955	acyltransferase
					STU_RS14585	hypothetical protein
					STU_RS14590	exopolysaccharide biosynthesis protein, glycosyltransferase
					STU_RS14595	hypothetical protein
					STU_RS18960	hypothetical protein
					STU_RS18965	glycosyltransferase family 2 protein
					STU_RS14605	polysaccharide pyruvyl transferase family protein
					STU_RS19480	hypothetical protein
					STU_RS14610	serine acetyltransferase
					STU_RS14615	multidrug MFS transporter
					STU_RS14620	UDP-N-acetylglucosamine--LPS N-acetylglucosamine transferase
6	4,591	1,328,986	1,333,577	32.8	STU_RS16390	hypothetical protein
					STU_RS16395	hypothetical protein
					STU_RS16400	transcriptional regulator
					STU_RS16405	hypothetical protein
					STU_RS16410	helix-turn-helix domain-containing protein
					STU_RS16415	radical SAM/SPASM domain-containing protein
7	10,513	1,467,717	1,478,230	40.5	STU_RS17135	hypothetical protein
					STU_RS17140	DNA polymerase IV
					STU_RS17145	formate C-acetyltransferase
					STU_RS17150	carbonate dehydratase
					STU_RS17155	GNAT family N-acetyltransferase
					STU_RS17160	pseudo
					STU_RS17165	type I restriction-modification system subunit S
					STU_RS18970	pseudo

					STU_RS17180	DNA recombinase
					STU_RS17185	hypothetical protein
					STU_RS17190	iron export ABC transporter permease subunit FetB
					STU_RS17195	putrescine/spermidine ABC transporter ATP-binding protein
					STU_RS17200	hypothetical protein
					STU_RS17205	serine hydrolase
8	13,572	1,585,316	1,598,888	38.1	STU_RS17705	type I glutamate--ammonia ligase
					STU_RS17710	MerR family transcriptional regulator
					STU_RS17715	aromatic acid exporter family protein
					STU_RS19735	pseudo
					STU_RS17725	phosphoglycerate kinase
					STU_RS17730	pseudo
					STU_RS19740	pseudo
					STU_RS17750	type I glyceraldehyde-3-phosphate dehydrogenase
					STU_RS17755	elongation factor G
					STU_RS17760	30S ribosomal protein S7
					STU_RS17765	30S ribosomal protein S12
					STU_RS17770	IS30 family transposase
					STU_RS17775	pur operon repressor
9	36,887	1,694,861	1,731,745	35.9	STU_RS18320	glutamate--cysteine ligase
					STU_RS18325	glutamate--cysteine ligase
					STU_RS18330	hypothetical protein
					STU_RS18335	helix-turn-helix domain-containing protein
					STU_RS18340	IS30 family transposase
					STU_RS18345	50S ribosomal protein L17
					STU_RS18350	DNA-directed RNA polymerase subunit alpha
					STU_RS18355	30S ribosomal protein S11
					STU_RS18360	30S ribosomal protein S13
					STU_RS19790	50S ribosomal protein L36
					STU_RS18365	translation initiation factor IF-1
					STU_RS18370	adenylate kinase
					STU_RS18375	preprotein translocase subunit SecY
					STU_RS18380	50S ribosomal protein L15
					STU_RS18385	50S ribosomal protein L30
					STU_RS18390	30S ribosomal protein S5
					STU_RS18395	50S ribosomal protein L18
					STU_RS18400	50S ribosomal protein L6
					STU_RS18405	30S ribosomal protein S8
					rpsN	30S ribosomal protein S14 type Z
					STU_RS18415	50S ribosomal protein L5
					STU_RS18420	50S ribosomal protein L24
					STU_RS18425	50S ribosomal protein L14
					STU_RS18430	30S ribosomal protein S17
					STU_RS18435	50S ribosomal protein L29
					STU_RS18440	50S ribosomal protein L16
					STU_RS18445	30S ribosomal protein S3
					STU_RS18450	50S ribosomal protein L22
					STU_RS18455	30S ribosomal protein S19
					STU_RS18460	50S ribosomal protein L2
					STU_RS18465	50S ribosomal protein L23
					STU_RS18470	50S ribosomal protein L4

					STU_RS18475	50S ribosomal protein L3
					STU_RS18480	30S ribosomal protein S10
					STU_RS18485	Holliday junction branch migration DNA helicase RuvB
					STU_RS18490	acetyltransferase
					STU_RS18495	membrane protein
					STU_RS18500	low molecular weight phosphotyrosine protein phosphatase
					STU_RS18505	adenylosuccinate synthetase
					STU_RS18510	DNA topology modulation protein FlaR
					STU_RS18515	hypothetical protein
					STU_RS18520	IS256 family transposase
					STU_RS18525	helix-turn-helix domain-containing protein
					STU_RS18530	hypothetical protein
					STU_RS18535	ABC transporter ATP-binding protein
					STU_RS18540	MutR family transcriptional regulator
					STU_RS18545	radical SAM protein
					STU_RS18550	ABC transporter ATP-binding protein
					STU_RS18560	replication initiation protein
					STU_RS18565	DUF3173 domain-containing protein
10	8,296	1,753,765	1,762,061	35.1	STU_RS18710	IS30 family transposase
					STU_RS18715	Asp23/Gls24 family envelope stress response protein
					STU_RS18720	CsbD family protein
					STU_RS18725	NUDIX domain-containing protein
					STU_RS19815	PadR family transcriptional regulator
					STU_RS19820	PadR family transcriptional regulator
					STU_RS18735	hypothetical protein
					STU_RS18740	hypothetical protein
					STU_RS18745	pseudo
					STU_RS18750	TetR/AcrR family transcriptional regulator

Strain MN-BM-A01						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	18,149	175,718	193,867	37.9	AMD33_RS00970	acyl carrier protein
					AMD33_RS00975	enoyl-[acyl-carrier-protein] reductase FabK
					AMD33_RS00980	[acyl-carrier-protein] S-malonyltransferase
					AMD33_RS00985	3-oxoacyl-[acyl-carrier-protein] reductase
					AMD33_RS00990	beta-ketoacyl-[acyl-carrier-protein] synthase II
					AMD33_RS00995	acetyl-CoA carboxylase biotin carboxyl carrier protein
					AMD33_RS01000	beta-hydroxyacyl-ACP dehydratase
					AMD33_RS01005	acetyl-CoA carboxylase biotin carboxylase subunit
					AMD33_RS01010	acetyl-CoA carboxylase carboxyltransferase subunit beta
					AMD33_RS01015	acetyl-CoA carboxylase carboxyl transferase subunit alpha
					AMD33_RS01020	S-ribosylhomocysteine lyase
					AMD33_RS09975	type I restriction endonuclease subunit S
					AMD33_RS09980	pseudo
					AMD33_RS01025	IS6 family transposase
					AMD33_RS01030	hypothetical protein
					AMD33_RS01035	DUF3114 domain-containing protein
					AMD33_RS01040	pseudo
					AMD33_RS01045	ABC transporter ATP-binding protein

					AMD33_RS01050	pseudo
					AMD33_RS01055	helix-turn-helix domain-containing protein
					AMD33_RS01060	pseudo
					AMD33_RS01065	IS3 family transposase
2	6,526	543,441	549,967	36.7	AMD33_RS02820	hypothetical protein
					AMD33_RS02825	virulence-associated protein E
					AMD33_RS02830	hypothetical protein
					AMD33_RS02835	hypothetical protein
					AMD33_RS02840	hypothetical protein
					AMD33_RS02845	hypothetical protein
					AMD33_RS02850	hypothetical protein
					AMD33_RS02855	hypothetical protein
					AMD33_RS02860	hypothetical protein
					AMD33_RS02865	XRE family transcriptional regulator
					AMD33_RS02870	XRE family transcriptional regulator
					AMD33_RS02875	site-specific integrase
3	23,694	614,875	638,569	35.2	AMD33_RS03220	UDP-galactopyranose mutase
					AMD33_RS03225	glycosyl transferase
					AMD33_RS03230	glycosyltransferase family 2 protein
					AMD33_RS03235	hypothetical protein
					AMD33_RS03240	hypothetical protein
					AMD33_RS03245	NAD(P)-dependent oxidoreductase
					AMD33_RS03250	pseudo
					AMD33_RS03255	glycosyltransferase family 2 protein
					AMD33_RS03260	glycosyltransferase family 2 protein
					AMD33_RS03265	hypothetical protein
					AMD33_RS03270	hypothetical protein
					AMD33_RS03275	pseudo
					AMD33_RS03280	hypothetical protein
					AMD33_RS10140	pseudo
					AMD33_RS03285	exopolysaccharide biosynthesis protein
					AMD33_RS03290	LytR family transcriptional regulator
					AMD33_RS10145	pseudo
					AMD33_RS03300	IS3 family transposase
					AMD33_RS03305	IS6-like element ISS1N family transposase
					AMD33_RS03310	VanZ family protein
					AMD33_RS03315	IS5-like element IS1194 family transposase
					AMD33_RS03320	pseudo
					AMD33_RS03325	hypothetical protein
					AMD33_RS10150	pseudo
					AMD33_RS03330	hypothetical protein
					AMD33_RS03335	glycosyltransferase family 2 protein
					AMD33_RS10155	pseudo
					AMD33_RS03350	IS3 family transposase
4	4,411	796,536	800,947	42.9	AMD33_RS04185	serine acetyltransferase
					AMD33_RS04190	cystathionine gamma-synthase
					AMD33_RS04195	cysteine synthase family protein
					AMD33_RS10310	hypothetical protein
					AMD33_RS04200	type I restriction endonuclease subunit R
5	8,420	1,186,024	1,194,444	32.9	AMD33_RS06150	hypothetical protein
					AMD33_RS06155	type II-A CRISPR-associated protein Csn2

					AMD33_RS06160	CRISPR-associated endonuclease Cas2
					AMD33_RS06165	type II CRISPR-associated endonuclease Cas1
					AMD33_RS06170	pseudo
6	10,188	1,443,892	1,454,080	39.3	AMD33_RS07425	phosphoglycerate kinase
					AMD33_RS07430	pseudo
					AMD33_RS10645	transposase
					AMD33_RS10650	transposase
					AMD33_RS07440	transposase
					AMD33_RS10655	pseudo
					AMD33_RS07455	type I glyceraldehyde-3-phosphate dehydrogenase
					AMD33_RS07460	elongation factor G
					AMD33_RS07465	30S ribosomal protein S7
					AMD33_RS07470	30S ribosomal protein S12
					AMD33_RS07475	ISL3 family transposase ISSth1
					AMD33_RS07480	pur operon repressor
7	48,298	1,511,272	1,559,570	38.0	AMD33_RS07845	prepilin-type N-terminal cleavage/methylation domain-containing protein
					AMD33_RS07850	competence protein CglB
					AMD33_RS07855	competence protein CglA
					AMD33_RS07860	DUF1033 domain-containing protein
					AMD33_RS07865	DNA-directed RNA polymerase subunit beta'
					AMD33_RS07870	DNA-directed RNA polymerase subunit beta
					AMD33_RS07875	pseudo
					AMD33_RS07880	pseudo
					AMD33_RS07885	tyrosine--tRNA ligase
					AMD33_RS07890	ketol-acid reductoisomerase
					AMD33_RS07895	acetolactate synthase small subunit
					AMD33_RS07900	acetolactate synthase, large subunit, biosynthetic type
					AMD33_RS07905	pseudo
					AMD33_RS07910	pseudo
					AMD33_RS07915	Asp23/Gls24 family envelope stress response protein
					AMD33_RS07920	MATE family efflux transporter
					AMD33_RS07925	threonine synthase
					AMD33_RS07930	aldehyde dehydrogenase
					AMD33_RS07935	aldehyde dehydrogenase
					AMD33_RS07940	alcohol dehydrogenase
					AMD33_RS07945	alcohol dehydrogenase
					AMD33_RS07950	aldehyde dehydrogenase
					AMD33_RS07955	endopeptidase
					AMD33_RS10690	pseudo
					AMD33_RS10695	pseudo
					AMD33_RS07980	GntR family transcriptional regulator
					AMD33_RS07985	UTRA domain-containing protein
					AMD33_RS07990	TIGR01440 family protein
					AMD33_RS07995	hypothetical protein
					AMD33_RS08000	hypothetical protein
					AMD33_RS08005	helix-turn-helix domain-containing protein
					AMD33_RS08010	IS30 family transposase
					AMD33_RS08015	ATP-binding protein
					AMD33_RS08020	dicarboxylate/amino acid:cation symporter
					AMD33_RS08025	alpha-L-glutamate ligase

					AMD33_RS08030	hypothetical protein
					AMD33_RS08035	hypothetical protein
					AMD33_RS08040	hypothetical protein
					AMD33_RS08045	hypothetical protein
					AMD33_RS08050	hypothetical protein
					AMD33_RS08055	DUF2513 domain-containing protein
					AMD33_RS08060	hypothetical protein
					AMD33_RS08065	hypothetical protein
					AMD33_RS08070	endonuclease
					AMD33_RS08075	XRE family transcriptional regulator
					AMD33_RS08080	DNA cytosine methyltransferase
					AMD33_RS08085	hypothetical protein
8	16,318	1,583,881	1,600,199	30.7	AMD33_RS08285	pseudo
					AMD33_RS08290	MFS transporter
					AMD33_RS08295	bacteriocin biosynthesis protein
					AMD33_RS08300	pseudo
					AMD33_RS08305	helix-turn-helix domain-containing protein
					AMD33_RS08310	hypothetical protein
					AMD33_RS08315	ABC transporter ATP-binding protein
					AMD33_RS08320	MutR family transcriptional regulator
					AMD33_RS08325	pseudo
					AMD33_RS08330	ABC transporter ATP-binding protein
					AMD33_RS08335	pseudo
					AMD33_RS08340	site-specific integrase
9	26,007	1,678,515	1,704,522	42.0	AMD33_RS08750	IS3 family transposase
					AMD33_RS08755	transposase
					AMD33_RS08760	transposase
					AMD33_RS08765	pseudo
					AMD33_RS08875	rod shape-determining protein MreC
					AMD33_RS08880	rod shape-determining protein MreD
					AMD33_RS08885	CHAP domain-containing protein
					AMD33_RS08890	ribose-phosphate diphosphokinase
					AMD33_RS10740	pseudo
					AMD33_RS08900	pyridoxal phosphate-dependent aminotransferase
					AMD33_RS08905	DNA repair protein RecO
					AMD33_RS08910	phosphate acyltransferase PlsX
					AMD33_RS08915	acyl carrier protein
					AMD33_RS08920	phosphoribosylaminoimidazolesuccinocarboxamide synthase
					AMD33_RS08925	phosphoribosylformylglycinamidine synthase
					AMD33_RS08930	amidophosphoribosyltransferase
					AMD33_RS08935	phosphoribosylformylglycinamidine cyclo-ligase
					AMD33_RS08940	phosphoribosylglycinamide formyltransferase
					AMD33_RS08945	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase PurH
10	7,918	1,767,123	1,775,041	30.1	AMD33_RS09315	pseudo
					AMD33_RS09320	lantibiotic biosynthesis protein
					AMD33_RS09325	lantibiotic transporter
					AMD33_RS09330	pseudo
					AMD33_RS09335	integrase
					AMD33_RS09340	Fis family transcriptional regulator

Strain MN-BM-A02						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	25,787	16,526	42,313	42.1	MNA02_RS00080	transposase
					MNA02_RS00085	IS3 family transposase
					MNA02_RS00195	rod shape-determining protein MreC
					MNA02_RS00200	rod shape-determining protein MreD
					MNA02_RS00205	CHAP domain-containing protein
					MNA02_RS00210	ribose-phosphate diphosphokinase
					MNA02_RS09775	pseudo
					MNA02_RS00220	pyridoxal phosphate-dependent aminotransferase
					MNA02_RS00225	DNA repair protein RecO
					MNA02_RS00230	phosphate acyltransferase PlsX
					MNA02_RS00235	acyl carrier protein
					MNA02_RS00240	phosphoribosylaminoimidazolesuccinocarboxamide synthase
					MNA02_RS00245	phosphoribosylformylglycinamidine synthase
					MNA02_RS00250	amidophosphoribosyltransferase
					MNA02_RS00255	phosphoribosylformylglycinamidine cyclo-ligase
					MNA02_RS00260	phosphoribosylglycinamide formyltransferase
					MNA02_RS00265	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/inosine monophosphate cyclohy
2	13,289	99,521	112,810	32.9	MNA02_RS00610	hypothetical protein
					MNA02_RS00615	2-dehydropantoate 2-reductase
					MNA02_RS00620	50S ribosomal protein L13
					MNA02_RS00625	30S ribosomal protein S9
					MNA02_RS00630	ISL3 family transposase
					MNA02_RS09805	pseudo
					MNA02_RS00635	helix-turn-helix transcriptional regulator
					MNA02_RS00640	hypothetical protein
					MNA02_RS00645	pseudo
					MNA02_RS00650	lantibiotic biosynthesis protein
					MNA02_RS00655	lantibiotic transporter
					MNA02_RS00660	pseudo
					MNA02_RS00665	pseudo
					MNA02_RS00670	DNA-binding protein
3	7,495	489,505	497,000	35.7	MNA02_RS02655	ABC transporter substrate-binding protein
					MNA02_RS09970	hypothetical protein
					MNA02_RS09975	pseudo
					MNA02_RS02665	macrolide ABC transporter
					MNA02_RS02670	ABC transporter permease, truncated
					MNA02_RS02675	50S ribosomal protein L10
					MNA02_RS02680	50S ribosomal protein L7/L12
					MNA02_RS02690	pseudo
					MNA02_RS09980	hypothetical protein
					MNA02_RS02695	ABC transporter ATP-binding protein
					MNA02_RS02700	ABC transporter permease
					MNA02_RS09985	hypothetical protein
					MNA02_RS02705	hypothetical protein
4	5,579	671,844	677,423	29.2	MNA02_RS03605	hypothetical protein
					MNA02_RS03610	hypothetical protein

					MNA02_RS03615	hypothetical protein
					MNA02_RS03620	restriction endonuclease subunit S
					MNA02_RS03625	NTPase
5	7,212	737,240	744,452	37.0	MNA02_RS03925	hypothetical protein
					MNA02_RS03930	hypothetical protein
					MNA02_RS03935	DNA primase
					MNA02_RS03940	hypothetical protein
					MNA02_RS03945	hypothetical protein
					MNA02_RS10090	pseudo
					MNA02_RS03955	hypothetical protein
					MNA02_RS03960	hypothetical protein
					MNA02_RS03965	hypothetical protein
					MNA02_RS03970	XRE family transcriptional regulator
					MNA02_RS10095	XRE family transcriptional regulator
					MNA02_RS03980	site-specific integrase
					MNA02_RS03985	hypothetical protein
6	14,545	800,097	814,642	38.4	MNA02_RS04275	IS256 family transposase
					MNA02_RS04280	copper-translocating P-type ATPase
					MNA02_RS04285	IS256 family transposase
					MNA02_RS04290	type I restriction endonuclease subunit R
					MNA02_RS10150	pseudo
					MNA02_RS04295	cysteine synthase family protein
					MNA02_RS04300	cystathionine gamma-synthase
					MNA02_RS10155	pseudo
					MNA02_RS10160	pseudo
					MNA02_RS04310	IS256 family transposase
					MNA02_RS10165	pseudo
					MNA02_RS04325	DNA methyltransferase
					MNA02_RS10170	pseudo
					MNA02_RS04330	sugar transporter
					MNA02_RS10175	pseudo
					MNA02_RS04335	hypothetical protein
7	6,466	854,409	860,875	31.1	MNA02_RS04550	MutR family transcriptional regulator
					MNA02_RS04555	Fe-S oxidoreductase
					MNA02_RS04560	IS256 family transposase
					MNA02_RS04565	radical SAM protein
					MNA02_RS04570	agmatinase
					MNA02_RS04575	MFS transporter
8	13,042	951,957	964,999	35.9	MNA02_RS05045	pseudo
					MNA02_RS05050	pseudo
					MNA02_RS05055	type I-E CRISPR-associated endoribonuclease Cas2
					MNA02_RS05060	type I-E CRISPR-associated endonuclease Cas1
					MNA02_RS05065	type I-E CRISPR-associated protein Cas6/Cse3/CasE
					MNA02_RS05070	type I-E CRISPR-associated protein Cas5/CasD
					MNA02_RS05075	type I-E CRISPR-associated protein Cas7/Cse4/CasC
					MNA02_RS05080	type I-E CRISPR-associated protein Cse2/CasB
					MNA02_RS05085	type I-E CRISPR-associated protein Cse1/CasA
					MNA02_RS05090	CRISPR-associated helicase/endonuclease Cas3
					MNA02_RS05095	DUF2207 domain-containing protein
9	5,607	1,201,043	1,206,650	26.4	MNA02_RS06325	MFS transporter permease
					MNA02_RS06330	oligoendopeptidase F

					MNA02_RS06335	helix-turn-helix domain-containing protein
					MNA02_RS06340	NUDIX hydrolase
					MNA02_RS06345	hypothetical protein
10	6,264	1,248,500	1,254,764	30.5	MNA02_RS06575	hypothetical protein
					MNA02_RS06580	hypothetical protein
					MNA02_RS06585	sodium transporter
					MNA02_RS06590	pseudo
					MNA02_RS06595	SPASM domain-containing protein
					MNA02_RS06600	transporter
					MNA02_RS06605	pseudo
					MNA02_RS10405	KxxxW-cyclized peptide pheromone
					MNA02_RS06610	helix-turn-helix domain-containing protein
11	4,070	1,735,404	1,739,474	38.4	MNA02_RS09065	type I restriction-modification system subunit M
					MNA02_RS09070	pseudo
					MNA02_RS10580	DNA-binding protein
12	39,601	1,745,108	1,784,706	35.7	MNA02_RS10590	helix-turn-helix domain-containing protein
					MNA02_RS09115	IS30 family transposase
					MNA02_RS09120	50S ribosomal protein L17
					MNA02_RS09125	DNA-directed RNA polymerase subunit alpha
					MNA02_RS09130	30S ribosomal protein S11
					MNA02_RS09135	30S ribosomal protein S13
					MNA02_RS10595	50S ribosomal protein L36
					MNA02_RS09140	translation initiation factor IF-1
					MNA02_RS09145	adenylate kinase
					MNA02_RS09150	preprotein translocase subunit SecY
					MNA02_RS09155	50S ribosomal protein L15
					MNA02_RS09160	50S ribosomal protein L30
					MNA02_RS09165	30S ribosomal protein S5
					MNA02_RS09170	50S ribosomal protein L18
					MNA02_RS09175	50S ribosomal protein L6
					MNA02_RS09180	30S ribosomal protein S8
					rpsN	30S ribosomal protein S14 type Z
					MNA02_RS09190	50S ribosomal protein L5
					MNA02_RS09195	50S ribosomal protein L24
					MNA02_RS09200	50S ribosomal protein L14
					MNA02_RS09205	30S ribosomal protein S17
					MNA02_RS09210	50S ribosomal protein L29
					MNA02_RS09215	50S ribosomal protein L16
					MNA02_RS09220	30S ribosomal protein S3
					MNA02_RS09225	50S ribosomal protein L22
					MNA02_RS09230	30S ribosomal protein S19
					MNA02_RS09235	50S ribosomal protein L2
					MNA02_RS09240	50S ribosomal protein L23
					MNA02_RS09245	50S ribosomal protein L4
					MNA02_RS09250	50S ribosomal protein L3
					MNA02_RS09255	30S ribosomal protein S10
					MNA02_RS09260	Holliday junction branch migration DNA helicase RuvB
					MNA02_RS09265	acyltransferase
					MNA02_RS09270	membrane protein
					MNA02_RS09275	low molecular weight phosphotyrosine protein phosphatase
					MNA02_RS09280	adenylosuccinate synthetase

					MNA02_RS09285	pseudo
					MNA02_RS09290	MFS transporter
					MNA02_RS09295	bacteriocin biosynthesis protein
					MNA02_RS09305	IS256 family transposase
					MNA02_RS09310	helix-turn-helix domain-containing protein
					MNA02_RS09315	serine/threonine protein kinase
					MNA02_RS09320	pseudo
					MNA02_RS09325	hypothetical protein
					MNA02_RS09330	ABC transporter ATP-binding protein
					MNA02_RS09335	MutR family transcriptional regulator
					MNA02_RS09340	radical SAM protein
					MNA02_RS09345	ABC transporter ATP-binding protein
					MNA02_RS09350	pseudo
					MNA02_RS10600	hypothetical protein
					MNA02_RS10605	replication initiator protein
					MNA02_RS09365	DUF3173 domain-containing protein
13	10,748	1,807,234	1,817,982	34.5	MNA02_RS09505	alkaline shock response membrane anchor protein AmaP
					MNA02_RS10625	DUF2273 domain-containing protein
					MNA02_RS09515	Asp23/Gls24 family envelope stress response protein
					MNA02_RS09520	CsbD family protein
					MNA02_RS09525	phosphohydrolase
					MNA02_RS09530	IS256 family transposase
					MNA02_RS10630	PadR family transcriptional regulator
					MNA02_RS10635	PadR family transcriptional regulator
					MNA02_RS09540	hypothetical protein
					MNA02_RS09545	hypothetical protein
					MNA02_RS09550	YhgE/Pip domain-containing protein
					MNA02_RS09555	TetR/AcrR family transcriptional regulator
					MNA02_RS09560	DUF1304 domain-containing protein
					MNA02_RS09565	30S ribosomal protein S4

Strain MN-ZLW-002						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	26,008	16,411	42,419	42.0	Y1U_RS00080	IS3 family transposase
					Y1U_RS09690	transposase
					Y1U_RS00085	transposase
					Y1U_RS00090	pseudo
					Y1U_RS00200	rod shape-determining protein MreC
					Y1U_RS00205	rod shape-determining protein MreD
					Y1U_RS00210	CHAP domain-containing protein
					Y1U_RS00215	ribose-phosphate diphosphokinase
					Y1U_RS09770	pseudo
					Y1U_RS00225	pyridoxal phosphate-dependent aminotransferase
					Y1U_RS00230	DNA repair protein RecO
					Y1U_RS00235	phosphate acyltransferase PlsX
					Y1U_RS00240	acyl carrier protein
					Y1U_RS00245	phosphoribosylaminoimidazolesuccinocarboxamide synthase
					Y1U_RS00250	phosphoribosylformylglycinamidine synthase
					Y1U_RS00255	amidophosphoribosyltransferase

					Y1U_RS00260	phosphoribosylformylglycinamidine cyclo-ligase
					Y1U_RS00265	phosphoribosylglycinamide formyltransferase
					Y1U_RS00270	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase PurH
2	7,919	105,020	112,939	30.1	Y1U_RS00635	pseudo
					Y1U_RS00640	lantibiotic biosynthesis protein
					Y1U_RS00645	lantibiotic transporter
					Y1U_RS00650	pseudo
					Y1U_RS00655	integrase
					Y1U_RS00660	Fis family transcriptional regulator
3	18,149	362,045	380,194	37.9	Y1U_RS02010	acyl carrier protein
					Y1U_RS02015	enoyl-[acyl-carrier-protein] reductase FabK
					Y1U_RS02020	[acyl-carrier-protein] S-malonyltransferase
					Y1U_RS02025	3-oxoacyl-[acyl-carrier-protein] reductase
					Y1U_RS02030	beta-ketoacyl-[acyl-carrier-protein] synthase II
					Y1U_RS02035	acetyl-CoA carboxylase biotin carboxyl carrier protein
					Y1U_RS02040	beta-hydroxyacyl-ACP dehydratase
					Y1U_RS02045	acetyl-CoA carboxylase biotin carboxylase subunit
					Y1U_RS02050	acetyl-CoA carboxylase carboxyltransferase subunit beta
					Y1U_RS02055	acetyl-CoA carboxylase carboxyl transferase subunit alpha
					Y1U_RS02060	S-ribosylhomocysteine lyase
					Y1U_RS09915	type I restriction endonuclease subunit S
					Y1U_RS09920	pseudo
					Y1U_RS02065	IS6 family transposase
					Y1U_RS02070	hypothetical protein
					Y1U_RS02075	DUF3114 domain-containing protein
					Y1U_RS02080	pseudo
					Y1U_RS02085	ABC transporter ATP-binding protein
					Y1U_RS02090	pseudo
					Y1U_RS02095	helix-turn-helix domain-containing protein
					Y1U_RS02100	pseudo
					Y1U_RS02105	IS3 family transposase
4	6,978	729,772	736,750	36.5	Y1U_RS03860	hypothetical protein
					Y1U_RS03865	virulence-associated protein E
					Y1U_RS03870	hypothetical protein
					Y1U_RS03875	hypothetical protein
					Y1U_RS03880	hypothetical protein
					Y1U_RS03885	hypothetical protein
					Y1U_RS03890	hypothetical protein
					Y1U_RS03895	hypothetical protein
					Y1U_RS03900	hypothetical protein
					Y1U_RS03905	XRE family transcriptional regulator
					Y1U_RS03910	XRE family transcriptional regulator
					Y1U_RS03915	site-specific integrase
					Y1U_RS03920	hypothetical protein
5	25,437	801,212	826,649	35.1	Y1U_RS04255	UDP-galactopyranose mutase
					Y1U_RS04260	glycosyl transferase
					Y1U_RS04265	glycosyltransferase family 2 protein
					Y1U_RS09715	hypothetical protein
					Y1U_RS09720	hypothetical protein
					Y1U_RS04275	NAD(P)-dependent oxidoreductase

					Y1U_RS04280	pseudo
					Y1U_RS04285	glycosyltransferase family 2 protein
					Y1U_RS04290	glycosyltransferase family 2 protein
					Y1U_RS04295	hypothetical protein
					Y1U_RS04300	hypothetical protein
					Y1U_RS04305	glycosyltransferase family 92 protein
					Y1U_RS04310	hypothetical protein
					Y1U_RS04315	pseudo
					Y1U_RS04320	exopolysaccharide biosynthesis protein
					Y1U_RS04325	LytR family transcriptional regulator
					Y1U_RS10090	pseudo
					Y1U_RS04335	IS3 family transposase
					Y1U_RS04340	IS6-like element ISS1N family transposase
					Y1U_RS04345	VanZ family protein
					Y1U_RS04350	IS5-like element IS1194 family transposase
					Y1U_RS09725	pseudo
					Y1U_RS04360	hypothetical protein
					Y1U_RS10095	pseudo
					Y1U_RS04365	glycosyltransferase family 2 protein
					Y1U_RS10100	pseudo
					Y1U_RS04375	IS3 family transposase
					Y1U_RS10105	pseudo
					Y1U_RS09730	phosphatase
6	4,411	982,889	987,300	42.9	Y1U_RS05200	serine acetyltransferase
					Y1U_RS05205	cystathionine gamma-synthase
					Y1U_RS05210	cysteine synthase family protein
					Y1U_RS10255	hypothetical protein
					Y1U_RS05215	type I restriction endonuclease subunit R
7	5,091	1,255,068	1,260,159	31.3	Y1U_RS06620	transporter
					Y1U_RS06625	KxxxW cyclic peptide radical SAM maturase
					Y1U_RS10445	KxxxW-cyclized peptide pheromone
					Y1U_RS06630	helix-turn-helix domain-containing protein
					Y1U_RS10450	pseudo
8	10,188	1,630,277	1,640,465	39.3	Y1U_RS08445	phosphoglycerate kinase
					Y1U_RS08450	pseudo
					Y1U_RS10575	transposase
					Y1U_RS10580	transposase
					Y1U_RS08460	transposase
					Y1U_RS10585	pseudo
					Y1U_RS08475	type I glyceraldehyde-3-phosphate dehydrogenase
					Y1U_RS08480	elongation factor G
					Y1U_RS08485	30S ribosomal protein S7
					Y1U_RS08490	30S ribosomal protein S12
					Y1U_RS08495	ISL3 family transposase ISSth1
					Y1U_RS08500	pur operon repressor
9	48,230	1,697,663	1,745,893	38.0	Y1U_RS08860	prepilin-type N-terminal cleavage/methylation domain-containing protein
					Y1U_RS08865	competence protein CglB
					Y1U_RS08870	competence protein CglA
					Y1U_RS08875	DUF1033 domain-containing protein
					Y1U_RS08880	DNA-directed RNA polymerase subunit beta'

					Y1U_RS08885	DNA-directed RNA polymerase subunit beta
					Y1U_RS08890	penicillin-binding protein
					Y1U_RS08895	tyrosine--tRNA ligase
					Y1U_RS08900	ketol-acid reductoisomerase
					Y1U_RS08905	acetolactate synthase small subunit
					Y1U_RS08910	acetolactate synthase, large subunit, biosynthetic type
					Y1U_RS08915	pseudo
					Y1U_RS08920	DAK2 domain-containing protein
					Y1U_RS08925	Asp23/Gls24 family envelope stress response protein
					Y1U_RS08930	MATE family efflux transporter
					Y1U_RS08935	threonine synthase
					Y1U_RS08940	aldehyde dehydrogenase
					Y1U_RS08945	aldehyde dehydrogenase
					Y1U_RS08950	alcohol dehydrogenase
					Y1U_RS08955	alcohol dehydrogenase
					Y1U_RS08960	aldehyde dehydrogenase
					Y1U_RS08965	endopeptidase
					Y1U_RS10625	pseudo
					Y1U_RS10630	pseudo
					Y1U_RS08990	GntR family transcriptional regulator
					Y1U_RS08995	UTRA domain-containing protein
					Y1U_RS09000	TIGR01440 family protein
					Y1U_RS09005	hypothetical protein
					Y1U_RS09010	hypothetical protein
					Y1U_RS09015	helix-turn-helix domain-containing protein
					Y1U_RS09020	IS30 family transposase
					Y1U_RS09025	ATP-binding protein
					Y1U_RS09030	dicarboxylate/amino acid:cation symporter
					Y1U_RS09035	alpha-L-glutamate ligase
					Y1U_RS09040	hypothetical protein
					Y1U_RS09045	hypothetical protein
					Y1U_RS09050	hypothetical protein
					Y1U_RS09055	hypothetical protein
					Y1U_RS09060	hypothetical protein
					Y1U_RS09065	DUF2513 domain-containing protein
					Y1U_RS09070	hypothetical protein
					Y1U_RS09075	hypothetical protein
					Y1U_RS09080	endonuclease
					Y1U_RS09085	XRE family transcriptional regulator
					Y1U_RS09090	DNA cytosine methyltransferase
					Y1U_RS09095	hypothetical protein
10	16,320	1,770,280	1,786,600	30.7	Y1U_RS09295	pseudo
					Y1U_RS09300	MFS transporter
					Y1U_RS09305	bacteriocin biosynthesis protein
					Y1U_RS09310	pseudo
					Y1U_RS09315	helix-turn-helix domain-containing protein
					Y1U_RS09320	hypothetical protein
					Y1U_RS09325	ABC transporter ATP-binding protein
					Y1U_RS09330	MutR family transcriptional regulator
					Y1U_RS09335	pseudo
					Y1U_RS09340	ABC transporter ATP-binding protein

					Y1U_RS09345	pseudo
					Y1U_RS09350	site-specific integrase

Strain NCTC12958						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	13,652	49,545	63,197	38.1	DQL34_RS00295	adenylosuccinate lyase
					DQL34_RS00300	arginine--tRNA ligase
					DQL34_RS00305	site-specific integrase
					DQL34_RS00310	pseudo
					DQL34_RS00315	transcriptional regulator
					DQL34_RS00320	hypothetical protein
					DQL34_RS00325	hypothetical protein
					DQL34_RS00330	hypothetical protein
					DQL34_RS00335	pseudo
					DQL34_RS00340	MerR family transcriptional regulator
					DQL34_RS00345	hypothetical protein
					DQL34_RS00350	DNA primase
					DQL34_RS00355	hypothetical protein
					DQL34_RS00360	pseudo
					DQL34_RS00365	hypothetical protein
					DQL34_RS00370	DUF3800 domain-containing protein
					DQL34_RS00375	pseudo
					argR	arginine repressor
2	6,696	113,975	120,671	30.7	DQL34_RS00690	hypothetical protein
					DQL34_RS00695	hypothetical protein
					DQL34_RS00700	XRE family transcriptional regulator
					DQL34_RS00705	pseudo
					DQL34_RS00710	branched-chain amino acid aminotransferase
					DQL34_RS00715	stage II sporulation protein M
					DQL34_RS00720	hypothetical protein
					DQL34_RS00725	ABC transporter ATP-binding protein
					DQL34_RS00730	ABC-2 transporter permease
					DQL34_RS00735	pseudo
					DQL34_RS00740	hypothetical protein
3	10,508	185,061	195,569	37.1	DQL34_RS01055	IS1182 family transposase
					DQL34_RS01060	toxin PezT
					DQL34_RS01065	DUF1837 domain-containing protein
					DQL34_RS01070	virulence associated protein
					DQL34_RS01075	hypothetical protein
					DQL34_RS01080	hypothetical protein
					DQL34_RS01085	plasmid mobilization relaxosome protein MobC
					DQL34_RS01090	relaxase
					DQL34_RS01095	hypothetical protein
4	5,047	208,223	213,270	38.9	DQL34_RS01160	hypothetical protein
					DQL34_RS01165	recombinase
					DQL34_RS01170	recombinase
					DQL34_RS01175	recombinase
					DQL34_RS01180	pseudo
5	8,104	500,114	508,218	42.1	DQL34_RS02695	DUF87 domain-containing protein

					DQL34_RS02700	CHAP domain-containing protein
					DQL34_RS02705	hypothetical protein
					DQL34_RS02710	hypothetical protein
					DQL34_RS02715	hypothetical protein
					DQL34_RS02720	LPXTG cell wall anchor domain-containing protein
					DQL34_RS02725	hypothetical protein
					DQL34_RS02730	type IV secretory system conjugative DNA transfer family protein
6	4,635	598,724	603,359	42.9	DQL34_RS03195	hypothetical protein
					DQL34_RS03200	transcriptional regulator
					DQL34_RS03205	CPBP family intramembrane metalloprotease
					DQL34_RS03210	hypothetical protein
					DQL34_RS03215	type IV secretory system conjugative DNA transfer family protein
					DQL34_RS03220	hypothetical protein
					DQL34_RS03225	conjugal transfer protein TrbL
					DQL34_RS03230	PrgI family protein
7	5,906	603,419	609,325	41.7	DQL34_RS03230	PrgI family protein
					DQL34_RS03235	AAA family ATPase
					DQL34_RS03240	pseudo
					DQL34_RS03245	transcriptional regulator
8	9,947	615,739	625,686	31.7	DQL34_RS03290	ATP-binding cassette domain-containing protein
					DQL34_RS03295	ATP-binding cassette domain-containing protein
					DQL34_RS03300	transporter
					DQL34_RS03305	ABC transporter permease
					DQL34_RS03310	helix-turn-helix domain-containing protein
					DQL34_RS03315	DUF3173 domain-containing protein
					DQL34_RS03320	site-specific integrase
					DQL34_RS03325	ISL3 family transposase
					DQL34_RS03330	hypothetical protein
					DQL34_RS03335	ABC transporter ATP-binding protein
9	38,494	726,278	764,772	40.5	DQL34_RS03885	phosphopyruvate hydratase
					DQL34_RS03890	site-specific integrase
					DQL34_RS03895	immunity protein
					DQL34_RS03900	ImmA/IrrE family metallo-endopeptidase
					DQL34_RS03905	XRE family transcriptional regulator
					DQL34_RS03910	XRE family transcriptional regulator
					DQL34_RS03915	DUF3102 domain-containing protein
					DQL34_RS03920	hypothetical protein
					DQL34_RS03925	excisionase
					DQL34_RS03930	hypothetical protein
					DQL34_RS03935	hypothetical protein
					DQL34_RS03940	DnaD domain protein
					DQL34_RS03945	DNA replication protein
					DQL34_RS03950	hypothetical protein
					DQL34_RS03955	single-stranded DNA-binding protein
					DQL34_RS03960	DUF1351 domain-containing protein
					DQL34_RS03965	single-stranded DNA-binding protein
					DQL34_RS03970	RusA family crossover junction endodeoxyribonuclease
					DQL34_RS03975	hypothetical protein
					DQL34_RS03980	hypothetical protein
					DQL34_RS03985	DUF1642 domain-containing protein

					DQL34_RS03990	DUF1642 domain-containing protein
					DQL34_RS03995	helix-turn-helix domain-containing protein
					DQL34_RS04000	DUF1372 family protein
					DQL34_RS04005	hypothetical protein
					DQL34_RS04010	hypothetical protein
					DQL34_RS04015	hypothetical protein
					DQL34_RS04020	DUF1340 domain-containing protein
					DQL34_RS04025	autolysin
					DQL34_RS04030	terminase small subunit
					DQL34_RS04035	PBSX family phage terminase large subunit
					DQL34_RS04040	phage portal protein
					DQL34_RS04045	hypothetical protein
					DQL34_RS04050	DUF4355 domain-containing protein
					DQL34_RS04055	hypothetical protein
					DQL34_RS04060	major capsid protein E
					DQL34_RS04065	hypothetical protein
					DQL34_RS04070	hypothetical protein
					DQL34_RS04075	histone H1
					DQL34_RS04080	DUF3168 domain-containing protein
					DQL34_RS04085	phage major tail protein, TP901-1 family
					DQL34_RS04090	hypothetical protein
					DQL34_RS04095	hypothetical protein
					DQL34_RS04100	phage tail protein
					DQL34_RS04105	phage tail protein
					DQL34_RS04110	hypothetical protein
					DQL34_RS04115	hypothetical protein
					DQL34_RS04120	DUF1366 domain-containing protein
					DQL34_RS04125	hypothetical protein
10	30,614	960,676	991,290	37.2	DQL34_RS05225	hypothetical protein
					DQL34_RS05230	hypothetical protein
					DQL34_RS05235	hypothetical protein
					DQL34_RS05240	hypothetical protein
					DQL34_RS05245	hypothetical protein
					DQL34_RS05250	hypothetical protein
					DQL34_RS05255	hypothetical protein
					DQL34_RS05260	DNA-entry nuclease
					DQL34_RS05265	hypothetical protein
					DQL34_RS05270	hypothetical protein
					DQL34_RS05275	hypothetical protein
					DQL34_RS05280	pseudo
					DQL34_RS05285	hypothetical protein
					DQL34_RS05290	type IA DNA topoisomerase
					DQL34_RS05295	AAA family ATPase
					DQL34_RS05300	hypothetical protein
					DQL34_RS05305	hypothetical protein
					DQL34_RS05310	LPXTG cell wall anchor domain-containing protein
					DQL34_RS05315	LPXTG cell wall anchor domain-containing protein
					DQL34_RS05320	single-stranded DNA-binding protein
					DQL34_RS05325	beta-hexosaminidase
					DQL34_RS05330	hypothetical protein
					DQL34_RS05335	hypothetical protein

					DQL34_RS05340	type IV secretory system conjugative DNA transfer family protein
11	5,103	1,047,931	1,053,034	30.3	DQL34_RS05645	hypothetical protein
					DQL34_RS05650	XRE family transcriptional regulator
					DQL34_RS05655	pseudo
					DQL34_RS05660	Fe-S oxidoreductase
					DQL34_RS05665	radical SAM protein
					DQL34_RS05670	arginase family protein
12	3,555	1,181,921	1,185,476	37.3	tnpA	IS200/IS605 family transposase
					DQL34_RS06360	50S ribosomal protein L20
					DQL34_RS06365	50S ribosomal protein L35
					DQL34_RS06370	translation initiation factor IF-3
					DQL34_RS06375	(d)CMP kinase
					DQL34_RS06380	LysM domain-containing protein
					DQL34_RS06385	ferredoxin
					DQL34_RS06390	EbsA protein
13	4,088	1,388,594	1,392,682	35.9	DQL34_RS07425	amino acid ABC transporter permease
					DQL34_RS07430	GNAT family N-acetyltransferase
					DQL34_RS07435	hypothetical protein
					DQL34_RS07440	aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme
					DQL34_RS07445	pseudo
14	8,397	1,429,880	1,438,277	37.3	DQL34_RS07655	pseudo
					DQL34_RS07660	HsdR family type I site-specific deoxyribonuclease
					DQL34_RS07665	restriction endonuclease subunit S
					DQL34_RS07670	type I restriction-modification system subunit M
15	16,226	1,466,413	1,482,639	37.8	DQL34_RS07790	hypothetical protein
					DQL34_RS07795	hypothetical protein
					DQL34_RS07800	pseudo
					DQL34_RS07805	ATP-binding cassette domain-containing protein
					DQL34_RS07810	pseudo
					DQL34_RS07815	pseudo
					DQL34_RS07820	bifunctional glutamate--cysteine ligase GshA/glutathione synthetase GshB
					DQL34_RS07825	hypothetical protein
					DQL34_RS07830	DUF1310 family protein
					DQL34_RS07835	pseudo
					DQL34_RS07840	ABC transporter ATP-binding protein
					DQL34_RS07845	pseudo
					DQL34_RS07850	IS256 family transposase
					DQL34_RS07855	helix-turn-helix domain-containing protein
					DQL34_RS07860	pseudo
					DQL34_RS07865	IS3 family transposase
16	6,025	1,720,856	1,726,881	41.0	DQL34_RS09115	pseudo
					DQL34_RS09120	GNAT family N-acetyltransferase
					DQL34_RS09125	pseudo
					DQL34_RS09130	pseudo
					DQL34_RS09135	pseudo
					DQL34_RS09140	pseudo
					DQL34_RS09145	hypothetical protein
					DQL34_RS09150	iron export ABC transporter permease subunit FetB
					DQL34_RS09155	ATP-binding cassette domain-containing protein

					DQL34_RS09160	hypothetical protein
					DQL34_RS09165	serine hydrolase
17	8,904	1,731,318	1,740,222	31.9	DQL34_RS09185	ArsR family transcriptional regulator
					DQL34_RS09190	EamA family transporter
					DQL34_RS09195	DUF4931 domain-containing protein
					DQL34_RS09200	hypothetical protein
					DQL34_RS09205	DlpA protein
					DQL34_RS09210	isocitrate/isopropylmalate dehydrogenase family protein
					DQL34_RS09215	alpha/beta hydrolase
					DQL34_RS09220	4-hydroxy-2-oxovalerate aldolase
					DQL34_RS09225	pseudo
18	7,228	1,746,276	1,753,504	36.5	DQL34_RS09255	hypothetical protein
					DQL34_RS09260	hypothetical protein
					DQL34_RS09265	ComC/BlpC family peptide pheromone/bacteriocin
					DQL34_RS09270	hypothetical protein
					DQL34_RS09275	ComC/BlpC family peptide pheromone/bacteriocin
					DQL34_RS09280	hypothetical protein
					DQL34_RS09285	hypothetical protein
					DQL34_RS09290	ComC/BlpC family peptide pheromone/bacteriocin
					DQL34_RS09295	hypothetical protein
					DQL34_RS09300	hypothetical protein
					DQL34_RS09305	ComC/BlpC family peptide pheromone/bacteriocin
					DQL34_RS09310	hypothetical protein
					DQL34_RS09315	hypothetical protein
					DQL34_RS09320	ComC/BlpC family peptide pheromone/bacteriocin
					DQL34_RS09325	DNA-binding response regulator
					DQL34_RS09330	GHKL domain-containing protein
19	23,933	1,840,027	1,863,960	38.9	DQL34_RS09720	hypothetical protein
					DQL34_RS09725	peptide transporter
					DQL34_RS09730	recombinase family protein
					DQL34_RS09735	relaxase
					DQL34_RS09740	plasmid mobilization relaxosome protein MobC
					DQL34_RS09745	transcriptional regulator
					DQL34_RS09750	DUF3991 domain-containing protein
					DQL34_RS09755	DNA topoisomerase III
					DQL34_RS09760	hypothetical protein
					DQL34_RS09765	type IV secretory system conjugative DNA transfer family protein
					DQL34_RS09770	hypothetical protein
					DQL34_RS09775	LPXTG cell wall anchor domain-containing protein
					DQL34_RS09780	hypothetical protein
					DQL34_RS09785	hypothetical protein
					DQL34_RS09790	hypothetical protein
					DQL34_RS09795	CHAP domain-containing protein
					DQL34_RS09800	DUF87 domain-containing protein
					DQL34_RS09805	hypothetical protein
					DQL34_RS09810	lactose transporter
					DQL34_RS09815	hypothetical protein
					DQL34_RS09820	hypothetical protein
					DQL34_RS09825	hypothetical protein
					DQL34_RS09830	hypothetical protein
					DQL34_RS09835	replication initiator protein A

					DQL34_RS09840	hypothetical protein
20	12,669	1,876,558	1,889,227	38.2	DQL34_RS09910	aromatic acid exporter family protein
					DQL34_RS09915	pseudo
					pgk	phosphoglycerate kinase
					tnpA	pseudo
					DQL34_RS09930	pseudo
					DQL34_RS09935	IS3 family transposase
					gap	type I glyceraldehyde-3-phosphate dehydrogenase
					fusA	elongation factor G
					DQL34_RS09950	30S ribosomal protein S7
					DQL34_RS09955	30S ribosomal protein S12
					DQL34_RS09960	pur operon repressor
					DQL34_RS09965	HD domain-containing protein
					DQL34_RS09970	DNA recombination protein RmuC
21	5,730	1,985,007	1,990,737	32.6	DQL34_RS10470	hypothetical protein
					DQL34_RS10475	pseudo
					DQL34_RS10480	helix-turn-helix domain-containing protein
					DQL34_RS10485	pseudo
					cadA	cadmium-translocating P-type ATPase
					lspA	signal peptidase II
					DQL34_RS10500	transcriptional regulator
					DQL34_RS10505	IS3 family transposase
22	17,111	2,000,607	2,017,718	39.7	DQL34_RS10570	hypothetical protein
					DQL34_RS10575	helix-turn-helix domain-containing protein
					DQL34_RS10580	IS30 family transposase
					DQL34_RS10585	50S ribosomal protein L17
					DQL34_RS10590	DNA-directed RNA polymerase subunit alpha
					DQL34_RS10595	30S ribosomal protein S11
					DQL34_RS10600	30S ribosomal protein S13
					DQL34_RS10605	50S ribosomal protein L36
					DQL34_RS10610	translation initiation factor IF-1
					DQL34_RS10615	adenylate kinase
					DQL34_RS10620	preprotein translocase subunit SecY
					DQL34_RS10625	50S ribosomal protein L15
					DQL34_RS10630	50S ribosomal protein L30
					DQL34_RS10635	30S ribosomal protein S5
					DQL34_RS10640	50S ribosomal protein L18
					DQL34_RS10645	50S ribosomal protein L6
					DQL34_RS10650	30S ribosomal protein S8
					DQL34_RS10655	type Z 30S ribosomal protein S14
					DQL34_RS10660	50S ribosomal protein L5
					DQL34_RS10665	50S ribosomal protein L24
					DQL34_RS10670	50S ribosomal protein L14
					DQL34_RS10675	30S ribosomal protein S17
					DQL34_RS10680	50S ribosomal protein L29
					DQL34_RS10685	50S ribosomal protein L16
					DQL34_RS10690	30S ribosomal protein S3
					DQL34_RS10695	50S ribosomal protein L22
					DQL34_RS10700	30S ribosomal protein S19
					DQL34_RS10705	50S ribosomal protein L2
					DQL34_RS10710	50S ribosomal protein L23

					DQL34_RS10715	50S ribosomal protein L4
					DQL34_RS10720	50S ribosomal protein L3
					DQL34_RS10725	30S ribosomal protein S10
					DQL34_RS10730	Holliday junction branch migration DNA helicase RuvB
23	13,427	2,024,288	2,037,715	33.4	DQL34_RS10765	MFS transporter
					DQL34_RS10770	ThiF family adenylyltransferase
					DQL34_RS10775	IS256 family transposase
					DQL34_RS10780	pseudo
					DQL34_RS10785	SDR family NAD(P)-dependent oxidoreductase
					DQL34_RS10790	DUF4649 family protein
					DQL34_RS10795	Rgg/GadR/MutR family transcriptional regulator
					DQL34_RS10800	radical SAM protein
					DQL34_RS10805	ABC transporter ATP-binding protein
					DQL34_RS10810	XRE family transcriptional regulator
					DQL34_RS10815	replication initiation protein
					DQL34_RS10820	DUF3173 domain-containing protein

Strain ND03						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	11,869	106,486	118,355	32.1	STND_RS00640	hypothetical protein
					STND_RS00645	ketopantoate reductase family protein
					STND_RS00650	50S ribosomal protein L13
					STND_RS00655	30S ribosomal protein S9
					STND_RS09780	pseudo
					STND_RS00660	helix-turn-helix transcriptional regulator
					STND_RS00665	hypothetical protein
					STND_RS00670	pseudo
					STND_RS00675	lantibiotic biosynthesis protein
					STND_RS00680	lantibiotic transporter
					STND_RS00685	pseudo
					STND_RS00690	integrase
					STND_RS00695	Fis family transcriptional regulator
2	22,592	367,071	389,663	38.4	STND_RS02045	acyl carrier protein
					STND_RS02050	enoyl-[acyl-carrier-protein] reductase FabK
					STND_RS02055	[acyl-carrier-protein] S-malonyltransferase
					STND_RS02060	3-oxoacyl-[acyl-carrier-protein] reductase
					STND_RS02065	beta-ketoacyl-[acyl-carrier-protein] synthase II
					STND_RS02070	acetyl-CoA carboxylase biotin carboxyl carrier protein
					STND_RS02075	beta-hydroxyacyl-ACP dehydratase
					STND_RS02080	acetyl-CoA carboxylase biotin carboxylase subunit
					STND_RS02085	acetyl-CoA carboxylase carboxyl transferase subunit beta
					STND_RS02090	acetyl-CoA carboxylase carboxyl transferase subunit alpha
					STND_RS02095	S-ribosylhomocysteine lyase
					STND_RS09900	pseudo
					STND_RS09905	type I restriction endonuclease
					STND_RS02100	pseudo
					STND_RS02105	glutamate decarboxylase
					STND_RS02110	amino acid permease
					STND_RS02115	hypothetical protein

					STND_RS02120	IS6 family transposase
					STND_RS02125	hypothetical protein
					STND_RS02130	DUF3114 domain-containing protein
					STND_RS02135	pseudo
					STND_RS02140	ABC transporter ATP-binding protein
					STND_RS02145	pseudo
					STND_RS02150	helix-turn-helix domain-containing protein
					STND_RS02155	pseudo
					STND_RS02160	IS3 family transposase
3	9,930	736,205	746,135	35.9	STND_RS03900	DNA-binding protein
					STND_RS03905	hypothetical protein
					STND_RS03910	hypothetical protein
					STND_RS03915	hypothetical protein
					STND_RS03920	DNA primase
					STND_RS03925	hypothetical protein
					STND_RS03930	hypothetical protein
					STND_RS03935	hypothetical protein
					STND_RS03940	hypothetical protein
					STND_RS03945	hypothetical protein
					STND_RS03950	hypothetical protein
					STND_RS03955	XRE family transcriptional regulator
					STND_RS10065	XRE family transcriptional regulator
					STND_RS03965	site-specific integrase
					STND_RS03970	hypothetical protein
4	18,845	786,036	804,881	37.6	STND_RS04205	ISL3 family transposase ISSth1
					STND_RS04210	cold-shock protein
					STND_RS04215	pseudo
					STND_RS04220	copper-translocating P-type ATPase
					STND_RS04225	pseudo
					STND_RS04230	pseudo
					STND_RS10105	hypothetical protein
					STND_RS04235	cysteine synthase family protein
					STND_RS04240	cystathionine gamma-synthase
					STND_RS04245	serine acetyltransferase
					STND_RS10110	pseudo
					STND_RS04250	IS256 family transposase
					STND_RS04255	sugar transporter
					STND_RS10115	pseudo
					STND_RS04260	hypothetical protein
					STND_RS04265	pseudo
					STND_RS04270	PrsW family intramembrane metalloprotease
					STND_RS04275	zinc ABC transporter substrate-binding protein AdcA
					STND_RS04280	hypothetical protein
5	5,144	840,524	845,668	29.3	STND_RS04460	MutR family transcriptional regulator
					STND_RS04465	Fe-S oxidoreductase
					STND_RS04470	radical SAM protein
					STND_RS04475	agmatinase
					STND_RS04480	MFS transporter
6	17,606	951,312	968,918	34.2	STND_RS05040	AI-2E family transporter
					STND_RS10220	pseudo
					STND_RS05050	pseudo

					STND_RS05055	pseudo
					STND_RS05060	phosphatase
					STND_RS05065	VanZ family protein
					STND_RS10225	hypothetical protein
					STND_RS10230	pseudo
					STND_RS05080	glycosyltransferase family 2 protein
					STND_RS10235	glycosyltransferase family 2 protein
					STND_RS05090	hypothetical protein
					STND_RS05095	glycosyl transferase
					STND_RS05100	multidrug MFS transporter
					STND_RS05105	UDP-N-acetylglucosamine--LPS N-acetylglucosamine transferase
					STND_RS05110	pseudo
					STND_RS05115	IS3 family transposase
					STND_RS10240	pseudo
					STND_RS10245	hypothetical protein
					STND_RS05130	pseudo
					STND_RS05135	flippase
					STND_RS10250	pseudo
					STND_RS05150	glycosyltransferase family 2 protein
					STND_RS09705	DUF1792 domain-containing protein
					STND_RS10255	DUF1792 domain-containing protein
7	5,091	1,239,425	1,244,516	31.3	STND_RS06580	transporter
					STND_RS06585	KxxxW cyclic peptide radical SAM maturase
					STND_RS10380	KxxxW-cyclized peptide pheromone
					STND_RS06590	helix-turn-helix domain-containing protein
					STND_RS10385	pseudo
8	14,915	1,722,030	1,736,945	31.3	STND_RS09005	hypothetical protein
					STND_RS09010	hypothetical protein
					STND_RS09015	pseudo
					STND_RS09020	ATP-binding protein
					STND_RS09025	dicarboxylate/amino acid:cation symporter
					STND_RS09030	alpha-L-glutamate ligase
					STND_RS09035	hypothetical protein
					STND_RS09040	hypothetical protein
					STND_RS09045	hypothetical protein
					STND_RS09050	hypothetical protein
					STND_RS09055	hypothetical protein
					STND_RS09060	DUF2513 domain-containing protein
					STND_RS09065	hypothetical protein
					STND_RS09070	hypothetical protein
					STND_RS09075	endonuclease
					STND_RS09080	XRE family transcriptional regulator
					STND_RS09085	DNA cytosine methyltransferase
					STND_RS09090	hypothetical protein
9	16,813	1,739,938	1,756,751	39.7	STND_RS10575	helix-turn-helix domain-containing protein
					STND_RS09120	IS30 family transposase
					STND_RS09125	50S ribosomal protein L17
					STND_RS09130	DNA-directed RNA polymerase subunit alpha
					STND_RS09135	30S ribosomal protein S11
					STND_RS09140	30S ribosomal protein S13
					STND_RS10580	50S ribosomal protein L36

					STND_RS09145	translation initiation factor IF-1
					STND_RS09150	adenylate kinase
					STND_RS09155	preprotein translocase subunit SecY
					STND_RS09160	50S ribosomal protein L15
					STND_RS09165	50S ribosomal protein L30
					STND_RS09170	30S ribosomal protein S5
					STND_RS09175	50S ribosomal protein L18
					STND_RS09180	50S ribosomal protein L6
					STND_RS09185	30S ribosomal protein S8
					rpsN	30S ribosomal protein S14 type Z
					STND_RS09195	50S ribosomal protein L5
					STND_RS09200	50S ribosomal protein L24
					STND_RS09205	50S ribosomal protein L14
					STND_RS09210	30S ribosomal protein S17
					STND_RS09215	50S ribosomal protein L29
					STND_RS09220	50S ribosomal protein L16
					STND_RS09225	30S ribosomal protein S3
					STND_RS09230	50S ribosomal protein L22
					STND_RS09235	30S ribosomal protein S19
					STND_RS09240	50S ribosomal protein L2
					STND_RS09245	50S ribosomal protein L23
					STND_RS09250	50S ribosomal protein L4
					STND_RS09255	50S ribosomal protein L3
					STND_RS09260	30S ribosomal protein S10
					STND_RS09265	Holliday junction branch migration DNA helicase RuvB
10	7,580	1,761,323	1,768,903	34.8	STND_RS09290	adenylate kinase
					STND_RS09295	MutR family transcriptional regulator
					STND_RS09300	radical SAM protein
					STND_RS09305	ABC transporter ATP-binding protein
					STND_RS09310	XRE family transcriptional regulator
					STND_RS10585	hypothetical protein
					STND_RS10590	replication initiator protein
					STND_RS09325	DUF3173 domain-containing protein

Strain ND07						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	6,761	68,588	75,349	37.3	BBD27_RS00360	site-specific integrase
					BBD27_RS09855	XRE family transcriptional regulator
					BBD27_RS00370	XRE family transcriptional regulator
					BBD27_RS00375	hypothetical protein
					BBD27_RS00380	hypothetical protein
					BBD27_RS00385	hypothetical protein
					BBD27_RS09860	pseudo
					BBD27_RS00390	hypothetical protein
					BBD27_RS00395	hypothetical protein
					BBD27_RS00400	DNA primase
					BBD27_RS00405	hypothetical protein
					BBD27_RS00410	hypothetical protein
2	5,579	135,166	140,745	29.2	BBD27_RS00705	NTPase

					BBD27_RS00710	restriction endonuclease subunit S
					BBD27_RS00715	hypothetical protein
					BBD27_RS00720	hypothetical protein
					BBD27_RS00725	hypothetical protein
3	15,535	311,070	326,605	37.0	BBD27_RS09940	hypothetical protein
					BBD27_RS01615	hypothetical protein
					BBD27_RS01620	hypothetical protein
					BBD27_RS01625	hypothetical protein
					BBD27_RS09945	hypothetical protein
					BBD27_RS01630	ABC transporter permease
					BBD27_RS09950	pseudo
					BBD27_RS01645	ABC transporter permease
					BBD27_RS01650	ABC transporter ATP-binding protein
					BBD27_RS09955	hypothetical protein
					BBD27_RS01660	pseudo
					BBD27_RS01670	50S ribosomal protein L7/L12
					BBD27_RS01675	50S ribosomal protein L10
					BBD27_RS01680	ABC transporter permease, truncated
					BBD27_RS01685	macrolide ABC transporter
					BBD27_RS09960	pseudo
					BBD27_RS09965	hypothetical protein
					BBD27_RS01695	ABC transporter substrate-binding protein
4	15,795	699,018	714,813	33.4	BBD27_RS03650	CapA family protein
					BBD27_RS03655	pseudo
					BBD27_RS03660	pseudo
					BBD27_RS03665	hypothetical protein
					BBD27_RS03670	DNA-binding protein
					BBD27_RS03675	pseudo
					BBD27_RS03680	pseudo
					BBD27_RS03685	lantibiotic transporter
					BBD27_RS03690	lantibiotic biosynthesis protein
					BBD27_RS03695	pseudo
					BBD27_RS03700	hypothetical protein
					BBD27_RS03705	helix-turn-helix transcriptional regulator
					BBD27_RS10120	pseudo
					BBD27_RS03710	ISL3 family transposase
					BBD27_RS03715	30S ribosomal protein S9
					BBD27_RS03720	50S ribosomal protein L13
5	15,941	783,680	799,621	42.3	BBD27_RS04105	phosphate acyltransferase PlsX
					BBD27_RS04110	DNA repair protein RecO
					BBD27_RS04115	pyridoxal phosphate-dependent aminotransferase
					BBD27_RS10150	pseudo
					BBD27_RS04125	ribose-phosphate diphosphokinase
					BBD27_RS04130	CHAP domain-containing protein
					BBD27_RS04135	rod shape-determining protein MreD
					BBD27_RS04140	rod shape-determining protein MreC
					BBD27_RS04250	IS3 family transposase
					BBD27_RS04255	transposase
6	9,984	848,600	858,584	34.8	BBD27_RS04490	30S ribosomal protein S4
					BBD27_RS04495	DUF1304 domain-containing protein
					BBD27_RS04500	TetR/AcrR family transcriptional regulator

					BBD27_RS04505	YhgE/Pip domain-containing protein
					BBD27_RS04510	hypothetical protein
					BBD27_RS04515	hypothetical protein
					BBD27_RS10160	PadR family transcriptional regulator
					BBD27_RS10165	PadR family transcriptional regulator
					BBD27_RS04525	IS256 family transposase
					BBD27_RS04530	phosphohydrolase
					BBD27_RS04535	CsbD family protein
					BBD27_RS04540	Asp23/Gls24 family envelope stress response protein
7	38,272	880,556	918,825	35.6	BBD27_RS04690	DUF3173 domain-containing protein
					BBD27_RS10175	replication initiator protein
					BBD27_RS10180	hypothetical protein
					BBD27_RS04705	pseudo
					BBD27_RS04710	ABC transporter ATP-binding protein
					BBD27_RS04715	radical SAM protein
					BBD27_RS04720	MutR family transcriptional regulator
					BBD27_RS04725	ABC transporter ATP-binding protein
					BBD27_RS04730	hypothetical protein
					BBD27_RS04735	helix-turn-helix domain-containing protein
					BBD27_RS04740	pseudo
					BBD27_RS04745	bacteriocin biosynthesis protein
					BBD27_RS04750	MFS transporter
					BBD27_RS04755	pseudo
					BBD27_RS04760	adenylosuccinate synthetase
					BBD27_RS04765	low molecular weight phosphotyrosine protein phosphatase
					BBD27_RS04770	membrane protein
					BBD27_RS04775	acyltransferase
					BBD27_RS04780	Holliday junction branch migration DNA helicase RuvB
					BBD27_RS04785	30S ribosomal protein S10
					BBD27_RS04790	50S ribosomal protein L3
					BBD27_RS04795	50S ribosomal protein L4
					BBD27_RS04800	50S ribosomal protein L23
					BBD27_RS04805	50S ribosomal protein L2
					BBD27_RS04810	30S ribosomal protein S19
					BBD27_RS04815	50S ribosomal protein L22
					BBD27_RS04820	30S ribosomal protein S3
					BBD27_RS04825	50S ribosomal protein L16
					BBD27_RS04830	50S ribosomal protein L29
					BBD27_RS04835	30S ribosomal protein S17
					BBD27_RS04840	50S ribosomal protein L14
					BBD27_RS04845	50S ribosomal protein L24
					BBD27_RS04850	50S ribosomal protein L5
					rpsN	30S ribosomal protein S14 type Z
					BBD27_RS04860	30S ribosomal protein S8
					BBD27_RS04865	50S ribosomal protein L6
					BBD27_RS04870	50S ribosomal protein L18
					BBD27_RS04875	30S ribosomal protein S5
					BBD27_RS04880	50S ribosomal protein L30
					BBD27_RS04885	50S ribosomal protein L15
					BBD27_RS04890	preprotein translocase subunit SecY
					BBD27_RS04895	adenylate kinase

					BBD27_RS04900	translation initiation factor IF-1
					BBD27_RS10185	50S ribosomal protein L36
					BBD27_RS04905	30S ribosomal protein S13
					BBD27_RS04910	30S ribosomal protein S11
					BBD27_RS04915	DNA-directed RNA polymerase subunit alpha
					BBD27_RS04920	50S ribosomal protein L17
					BBD27_RS04925	IS30 family transposase
					BBD27_RS10190	helix-turn-helix domain-containing protein
8	4,070	924,460	928,530	38.4	BBD27_RS04965	DNA-binding protein
					BBD27_RS04970	pseudo
					BBD27_RS04975	type I restriction-modification system subunit M
9	7,681	1,409,172	1,416,853	30.7	BBD27_RS07415	helix-turn-helix domain-containing protein
					BBD27_RS10370	KxxxW-cyclized peptide pheromone
					BBD27_RS07420	pseudo
					BBD27_RS07425	transporter
					BBD27_RS07430	SPASM domain-containing protein
					BBD27_RS07435	pseudo
					BBD27_RS07440	sodium transporter
					BBD27_RS07445	hypothetical protein
					BBD27_RS07450	hypothetical protein
					BBD27_RS07455	ABC transporter ATP-binding protein
10	4,400	1,457,288	1,461,688	26.1	BBD27_RS07675	hypothetical protein
					BBD27_RS07680	NUDIX hydrolase
					BBD27_RS07685	helix-turn-helix domain-containing protein
					BBD27_RS07690	oligoendopeptidase F
					BBD27_RS07695	MFS transporter permease
11	13,042	1,698,948	1,711,990	35.9	BBD27_RS08925	DUF2207 domain-containing protein
					BBD27_RS08930	CRISPR-associated helicase/endonuclease Cas3
					BBD27_RS08935	type I-E CRISPR-associated protein Cse1/CasA
					BBD27_RS08940	type I-E CRISPR-associated protein Cse2/CasB
					BBD27_RS08945	type I-E CRISPR-associated protein Cas7/Cse4/CasC
					BBD27_RS08950	type I-E CRISPR-associated protein Cas5/CasD
					BBD27_RS08955	type I-E CRISPR-associated protein Cas6/Cse3/CasE
					BBD27_RS08960	type I-E CRISPR-associated endonuclease Cas1
					BBD27_RS08965	type I-E CRISPR-associated endoribonuclease Cas2
					BBD27_RS08970	pseudo
					BBD27_RS08975	pseudo
12	16,314	1,843,953	1,860,267	37.6	BBD27_RS09660	putative sulfate exporter family transporter
					BBD27_RS09665	hypothetical protein
					BBD27_RS09670	hypothetical protein
					BBD27_RS09675	pseudo
					BBD27_RS09680	PrsW family intramembrane metalloprotease
					BBD27_RS09685	pseudo
					BBD27_RS09690	hypothetical protein
					BBD27_RS10605	pseudo
					BBD27_RS09695	sugar transporter
					BBD27_RS10610	pseudo
					BBD27_RS09700	DNA methyltransferase
					BBD27_RS10615	pseudo
					BBD27_RS09715	pseudo
					BBD27_RS10620	pseudo

					BBD27_RS09720	pseudo
					BBD27_RS09725	cystathionine gamma-synthase
					BBD27_RS09730	cysteine synthase family protein
					BBD27_RS10625	pseudo
					BBD27_RS09735	type I restriction endonuclease subunit R
					BBD27_RS09740	IS256 family transposase

Strain S9						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	6,166	94,597	100,763	29.9	AVT04_RS00500	MFS transporter
					AVT04_RS00505	agmatinase
					AVT04_RS00510	radical SAM protein
					AVT04_RS00515	Fe-S oxidoreductase
					AVT04_RS00520	MutR family transcriptional regulator
					AVT04_RS00525	XRE family transcriptional regulator
					AVT04_RS00530	pseudo
2	4,109	126,095	130,204	40.3	AVT04_RS00655	site-specific DNA-methyltransferase
					AVT04_RS00660	type III restriction endonuclease
					AVT04_RS00665	pseudo
3	12,538	149,021	161,559	38.7	AVT04_RS00785	sugar transporter
					AVT04_RS00790	pseudo
					AVT04_RS00795	pseudo
					AVT04_RS09595	pseudo
					AVT04_RS09600	pseudo
					AVT04_RS00810	IS256 family transposase
					AVT04_RS09605	pseudo
					AVT04_RS00815	pseudo
					AVT04_RS00820	cystathionine gamma-synthase
					AVT04_RS00825	cysteine synthase family protein
					AVT04_RS09610	hypothetical protein
					AVT04_RS00830	type I restriction endonuclease subunit R
					AVT04_RS00835	IS256 family transposase
					AVT04_RS00840	copper-translocating P-type ATPase
4	4,695	263,339	268,034	28.2	AVT04_RS01390	restriction endonuclease subunit S
					AVT04_RS01395	hypothetical protein
					AVT04_RS01400	hypothetical protein
5	4,411	283,671	288,082	26.5	AVT04_RS01480	hypothetical protein
					AVT04_RS01485	XRE family transcriptional regulator
					AVT04_RS01490	hypothetical protein
					AVT04_RS01495	hypothetical protein
					AVT04_RS01500	hypothetical protein
					AVT04_RS01505	hypothetical protein
					AVT04_RS01510	hypothetical protein
					AVT04_RS01515	hypothetical protein
6	19,736	813,720	833,456	34.0	AVT04_RS04305	ACT domain-containing protein
					AVT04_RS04310	hypothetical protein
					AVT04_RS04315	pseudo
					AVT04_RS04325	pseudo
					AVT04_RS04330	pseudo

					AVT04_RS04335	IS30 family transposase
					AVT04_RS04340	pseudo
					AVT04_RS04345	pseudo
					AVT04_RS04350	pseudo
					AVT04_RS04355	pseudo
					AVT04_RS04360	lantibiotic transporter
					AVT04_RS04365	lantibiotic biosynthesis protein
					AVT04_RS04370	pseudo
					AVT04_RS04375	hypothetical protein
					AVT04_RS04380	pseudo
					AVT04_RS09860	pseudo
					AVT04_RS09865	hypothetical protein
					AVT04_RS04385	30S ribosomal protein S9
					AVT04_RS04390	50S ribosomal protein L13
					AVT04_RS04395	2-dehydropantoate 2-reductase
					AVT04_RS04400	hypothetical protein
					AVT04_RS09870	transcriptional regulator
					AVT04_RS04405	DegV family protein
					AVT04_RS04410	hypothetical protein
7	15,951	994,003	1,009,954	30.8	AVT04_RS05340	DUF3173 domain-containing protein
					AVT04_RS05345	replication initiation protein
					AVT04_RS05355	ABC transporter ATP-binding protein
					AVT04_RS05360	radical SAM protein
					AVT04_RS05365	MutR family transcriptional regulator
					AVT04_RS05370	ABC transporter ATP-binding protein
					AVT04_RS05375	hypothetical protein
					AVT04_RS05380	helix-turn-helix domain-containing protein
					AVT04_RS05385	pseudo
					AVT04_RS05390	MccB-like protein
					AVT04_RS05395	MFS transporter
					AVT04_RS05400	DNA topology modulation protein FlaR
8	21,101	1,226,895	1,247,996	39.4	AVT04_RS10015	hypothetical protein
					AVT04_RS06610	peptide cleavage/export ABC transporter
					AVT04_RS06615	pseudo
					AVT04_RS06620	ComC/BlpC family peptide pheromone/bacteriocin
					AVT04_RS06625	ATP-binding protein
					AVT04_RS06630	DNA-binding response regulator
					AVT04_RS06635	ComC/BlpC family peptide pheromone/bacteriocin
					AVT04_RS10020	pseudo
					AVT04_RS06650	hypothetical protein
					AVT04_RS06655	Xaa-Pro dipeptidyl-peptidase
					AVT04_RS06660	aquaporin family protein
					AVT04_RS06665	CPBP family intramembrane metalloprotease
					AVT04_RS06670	IS30 family transposase
					AVT04_RS06675	peptidase
					AVT04_RS06680	serine hydrolase
					AVT04_RS06685	hypothetical protein
					AVT04_RS06690	ABC transporter ATPase
					AVT04_RS06695	iron export ABC transporter permease subunit FetB
					AVT04_RS06700	hypothetical protein
					AVT04_RS06705	DNA recombinase

					AVT04_RS06710	recombinase
					AVT04_RS06715	plasmid mobilization protein
					AVT04_RS06720	type I restriction-modification system subunit S
					AVT04_RS10025	pseudo
					AVT04_RS06725	pseudo
9	7,880	1,526,859	1,534,739	35.5	AVT04_RS08110	amidohydrolase
					AVT04_RS08115	hypothetical protein
					AVT04_RS08120	glycerol dehydrogenase
					AVT04_RS08125	glycerol dehydrogenase
					AVT04_RS08130	aminotransferase
					AVT04_RS08135	hypothetical protein
					AVT04_RS08140	GNAT family N-acetyltransferase
					AVT04_RS08145	amino acid ABC transporter permease
					AVT04_RS08150	amino acid ABC transporter ATP-binding protein
					AVT04_RS08155	adenylosuccinate lyase
					AVT04_RS08160	amino acid ABC transporter substrate-binding protein
10	10,035	1,750,108	1,760,143	33.0	AVT04_RS09285	glycosyltransferase family 1 protein
					AVT04_RS09290	acyltransferase
					AVT04_RS09295	glycosyltransferase family 2 protein
					AVT04_RS09300	EpsG family protein
					AVT04_RS09305	exopolysaccharide gene cluster protein
					AVT04_RS09310	exopolysaccharide polymerization protein
					AVT04_RS09315	exopolysaccharide biosynthesis protein
					AVT04_RS09320	VanZ family protein
					AVT04_RS09325	pseudo
					AVT04_RS09330	lipopolysaccharide 1,6-galactosyltransferase

Strain SMQ-301						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	28,370	488,729	517,099	39.5	SMQ301_RS02675	hypothetical protein
					SMQ301_RS10650	pseudo
					SMQ301_RS02680	PTS beta-glucoside transporter subunit IIABC
					SMQ301_RS09950	pseudo
					SMQ301_RS02690	peptidylprolyl isomerase
					SMQ301_RS10655	pseudo
					SMQ301_RS02705	hypothetical protein
					SMQ301_RS02710	hypothetical protein
					SMQ301_RS02715	hypothetical protein
					SMQ301_RS02720	hypothetical protein
					SMQ301_RS02725	LysR family transcriptional regulator
					SMQ301_RS02730	signal peptidase II
					SMQ301_RS02735	RluA family pseudouridine synthase
					SMQ301_RS02740	bifunctional pyrimidine operon transcriptional regulator/uracil phosphoribosyltransferase
					SMQ301_RS02745	uracil transporter
					SMQ301_RS02750	aspartate carbamoyltransferase catalytic subunit
					SMQ301_RS02755	carbamoyl-phosphate synthase small subunit
					SMQ301_RS02760	carbamoyl-phosphate synthase large chain
					SMQ301_RS02765	ISL3 family transposase

					SMQ301_RS02770	SAM-dependent methyltransferase
					SMQ301_RS10660	pseudo
					SMQ301_RS02775	efflux RND transporter periplasmic adaptor subunit
					SMQ301_RS09955	pseudo
					SMQ301_RS02785	50S ribosomal protein L10
					SMQ301_RS02790	50S ribosomal protein L7/L12
					SMQ301_RS02795	ISL3 family transposase
					SMQ301_RS10665	hypothetical protein
					SMQ301_RS02800	ABC transporter ATP-binding protein
					SMQ301_RS02805	ABC transporter permease
					SMQ301_RS10670	pseudo
					SMQ301_RS02810	hypothetical protein
2	8,127	752,801	760,928	36.6	SMQ301_RS04010	hypothetical protein
					SMQ301_RS04015	hypothetical protein
					SMQ301_RS04020	hypothetical protein
					SMQ301_RS04025	hypothetical protein
					SMQ301_RS04030	hypothetical protein
					SMQ301_RS04035	DNA primase
					SMQ301_RS04040	hypothetical protein
					SMQ301_RS04045	hypothetical protein
					SMQ301_RS04050	hypothetical protein
					SMQ301_RS04055	hypothetical protein
					SMQ301_RS04060	hypothetical protein
					SMQ301_RS10030	pseudo
					SMQ301_RS04070	XRE family transcriptional regulator
					SMQ301_RS04075	XRE family transcriptional regulator
3	10,071	817,873	827,944	39.2	SMQ301_RS04380	copper-translocating P-type ATPase
					SMQ301_RS04385	IS256 family transposase
					SMQ301_RS10055	pseudo
					SMQ301_RS10840	hypothetical protein
					SMQ301_RS04395	cysteine synthase family protein
					SMQ301_RS04400	cystathionine gamma-synthase
					SMQ301_RS04405	serine acetyltransferase
					SMQ301_RS10845	pseudo
					SMQ301_RS04410	IS256 family transposase
					SMQ301_RS10850	hypothetical protein
4	5,144	865,794	870,938	29.3	SMQ301_RS04600	MutR family transcriptional regulator
					SMQ301_RS04605	Fe-S oxidoreductase
					SMQ301_RS04610	radical SAM protein
					SMQ301_RS04615	agmatinase
					SMQ301_RS04620	MFS transporter
5	4,426	1,052,393	1,056,819	38.9	SMQ301_RS05580	DUF3492 domain-containing protein
					SMQ301_RS05585	DUF2194 domain-containing protein
					SMQ301_RS05590	hypothetical protein
					SMQ301_RS05595	membrane protein
6	5,607	1,215,056	1,220,663	26.4	SMQ301_RS06405	MFS transporter permease
					SMQ301_RS06410	oligoendopeptidase F
					SMQ301_RS06415	helix-turn-helix domain-containing protein
					SMQ301_RS10230	NUDIX hydrolase
					SMQ301_RS06425	hypothetical protein
7	9,259	1,261,526	1,270,785	31.5	SMQ301_RS06640	ABC transporter ATP-binding protein

					SMQ301_RS06645	hypothetical protein
					SMQ301_RS06650	hypothetical protein
					SMQ301_RS06655	sodium transporter
					SMQ301_RS10250	pseudo
					SMQ301_RS10255	SPASM domain-containing protein
					SMQ301_RS06670	transporter
					SMQ301_RS06675	KxxxW cyclic peptide radical SAM maturase
					SMQ301_RS11110	KxxxW-cyclized peptide pheromone
					SMQ301_RS06680	helix-turn-helix domain-containing protein
					SMQ301_RS11115	pseudo
8	4,495	1,540,485	1,544,980	36.9	SMQ301_RS08030	hypothetical protein
					SMQ301_RS08035	hypothetical protein
					SMQ301_RS08040	ComC/BlpC family peptide pheromone/bacteriocin
					SMQ301_RS08045	hypothetical protein
					SMQ301_RS11225	hypothetical protein
					SMQ301_RS08055	ComC/BlpC family peptide pheromone/bacteriocin
					SMQ301_RS11230	hypothetical protein
					SMQ301_RS08060	hypothetical protein
					SMQ301_RS08065	ComC/BlpC family peptide pheromone/bacteriocin
					SMQ301_RS08070	DNA-binding response regulator
					SMQ301_RS08075	GHKL domain-containing protein
9	38,272	1,759,603	1,797,872	35.6	SMQ301_RS11310	helix-turn-helix domain-containing protein
					SMQ301_RS09200	IS30 family transposase
					SMQ301_RS09205	50S ribosomal protein L17
					SMQ301_RS09210	DNA-directed RNA polymerase subunit alpha
					SMQ301_RS09215	30S ribosomal protein S11
					SMQ301_RS09220	30S ribosomal protein S13
					SMQ301_RS11315	50S ribosomal protein L36
					SMQ301_RS09225	translation initiation factor IF-1
					SMQ301_RS09230	adenylate kinase
					SMQ301_RS09235	preprotein translocase subunit SecY
					SMQ301_RS09240	50S ribosomal protein L15
					SMQ301_RS09245	50S ribosomal protein L30
					SMQ301_RS09250	30S ribosomal protein S5
					SMQ301_RS09255	50S ribosomal protein L18
					SMQ301_RS09260	50S ribosomal protein L6
					SMQ301_RS09265	30S ribosomal protein S8
					rpsN	30S ribosomal protein S14 type Z
					SMQ301_RS09275	50S ribosomal protein L5
					SMQ301_RS09280	50S ribosomal protein L24
					SMQ301_RS09285	50S ribosomal protein L14
					SMQ301_RS09290	30S ribosomal protein S17
					SMQ301_RS09295	50S ribosomal protein L29
					SMQ301_RS09300	50S ribosomal protein L16
					SMQ301_RS09305	30S ribosomal protein S3
					SMQ301_RS09310	50S ribosomal protein L22
					SMQ301_RS09315	30S ribosomal protein S19
					SMQ301_RS09320	50S ribosomal protein L2
					SMQ301_RS09325	50S ribosomal protein L23
					SMQ301_RS09330	50S ribosomal protein L4
					SMQ301_RS09335	50S ribosomal protein L3

					SMQ301_RS09340	30S ribosomal protein S10
					SMQ301_RS09345	Holliday junction branch migration DNA helicase RuvB
					SMQ301_RS09350	acetyltransferase
					SMQ301_RS09355	membrane protein
					SMQ301_RS09360	low molecular weight phosphotyrosine protein phosphatase
					SMQ301_RS09365	adenylosuccinate synthetase
					SMQ301_RS10370	pseudo
					SMQ301_RS09375	MFS transporter
					SMQ301_RS09380	bacteriocin biosynthesis protein
					SMQ301_RS10375	pseudo
					SMQ301_RS09395	helix-turn-helix domain-containing protein
					SMQ301_RS09400	hypothetical protein
					SMQ301_RS09405	ABC transporter ATP-binding protein
					SMQ301_RS09410	MutR family transcriptional regulator
					SMQ301_RS09415	radical SAM protein
					SMQ301_RS09420	ABC transporter ATP-binding protein
					SMQ301_RS09425	XRE family transcriptional regulator
					SMQ301_RS11320	hypothetical protein
					SMQ301_RS11325	replication initiator protein
					SMQ301_RS09440	DUF3173 domain-containing protein

Strain ST3						
GI	**Size**	**Genomic coordinates**		**GC (%)**	**Locus_tag**	**Annotated function**
1	6,219	510,480	516,699	34.6	BGL51_RS02745	50S ribosomal protein L7/L12
					BGL51_RS02750	pseudo
					BGL51_RS02755	hypothetical protein
					BGL51_RS02760	ABC transporter ATP-binding protein
					BGL51_RS02765	ABC transporter permease
					BGL51_RS02770	hypothetical protein
					BGL51_RS02775	hypothetical protein
					BGL51_RS02780	hypothetical protein
2	9,374	753,973	763,347	36.7	BGL51_RS03980	hypothetical protein
					BGL51_RS03985	hypothetical protein
					BGL51_RS03990	hypothetical protein
					BGL51_RS03995	hypothetical protein
					BGL51_RS04000	hypothetical protein
					BGL51_RS04005	DNA primase
					BGL51_RS04010	hypothetical protein
					BGL51_RS04015	hypothetical protein
					BGL51_RS04020	hypothetical protein
					BGL51_RS04025	hypothetical protein
					BGL51_RS04030	pseudo
					BGL51_RS04035	pseudo
					BGL51_RS04040	XRE family transcriptional regulator
					BGL51_RS04045	XRE family transcriptional regulator
					BGL51_RS04050	site-specific integrase
3	20,237	826,129	846,366	37.8	BGL51_RS04400	cold-shock protein
					BGL51_RS04405	ISL3-like element ISSth1 family transposase
					BGL51_RS04410	pseudo

					BGL51_RS04415	copper-translocating P-type ATPase
					BGL51_RS04420	pseudo
					BGL51_RS04425	pseudo
					BGL51_RS04430	hypothetical protein
					BGL51_RS04435	cysteine synthase family protein
					BGL51_RS04440	cystathionine gamma-synthase
					BGL51_RS04445	pseudo
					BGL51_RS04450	pseudo
					BGL51_RS04455	pseudo
					BGL51_RS04460	pseudo
					BGL51_RS04465	DNA methyltransferase
					BGL51_RS04470	pseudo
					BGL51_RS04475	sugar transporter
					BGL51_RS04480	pseudo
					BGL51_RS04485	type I restriction-modification system subunit R
					BGL51_RS04490	pseudo
					BGL51_RS04495	PrsW family intramembrane metalloprotease
					BGL51_RS04500	zinc ABC transporter substrate-binding protein AdcA
					BGL51_RS04505	hypothetical protein
4	5,607	1,224,942	1,230,549	26.4	BGL51_RS06610	MFS transporter permease
					BGL51_RS06615	oligoendopeptidase F
					BGL51_RS06620	helix-turn-helix domain-containing protein
					BGL51_RS06625	NUDIX hydrolase
					BGL51_RS06630	hypothetical protein
5	6,224	1,266,908	1,273,132	34.2	BGL51_RS06815	pseudo
					BGL51_RS06820	hypothetical protein
					BGL51_RS06825	DNA-directed DNA polymerase
					BGL51_RS06830	pseudo
					BGL51_RS06835	hypothetical protein
					BGL51_RS06840	hypothetical protein
					BGL51_RS06845	LysE family translocator
					BGL51_RS06850	hypothetical protein
					BGL51_RS06855	pseudo
6	10,748	1,749,066	1,759,814	33.0	BGL51_RS09220	helix-turn-helix domain-containing protein
					BGL51_RS09225	IS30 family transposase
					BGL51_RS09230	ATP-binding protein
					BGL51_RS09235	dicarboxylate/amino acid:cation symporter
					BGL51_RS09240	alpha-L-glutamate ligase
					BGL51_RS09245	ISL3-like element ISSth1 family transposase
					BGL51_RS09250	hypothetical protein
					BGL51_RS09255	pseudo
					BGL51_RS09260	hypothetical protein
					BGL51_RS09265	CHAP domain-containing protein
					BGL51_RS09270	DNA-binding protein
					BGL51_RS09275	site-specific integrase
7	39,428	1,762,645	1,802,070	35.6	BGL51_RS09305	helix-turn-helix domain-containing protein
					BGL51_RS09310	hypothetical protein
					BGL51_RS09315	50S ribosomal protein L17
					BGL51_RS09320	DNA-directed RNA polymerase subunit alpha
					BGL51_RS09325	30S ribosomal protein S11
					BGL51_RS09330	30S ribosomal protein S13

					BGL51_RS09335	50S ribosomal protein L36
					BGL51_RS09340	translation initiation factor IF-1
					BGL51_RS09345	adenylate kinase
					BGL51_RS09350	preprotein translocase subunit SecY
					BGL51_RS09355	50S ribosomal protein L15
					BGL51_RS09360	50S ribosomal protein L30
					BGL51_RS09365	30S ribosomal protein S5
					BGL51_RS09370	50S ribosomal protein L18
					BGL51_RS09375	50S ribosomal protein L6
					BGL51_RS09380	30S ribosomal protein S8
					BGL51_RS09385	30S ribosomal protein S14 type Z
					BGL51_RS09390	50S ribosomal protein L5
					BGL51_RS09395	50S ribosomal protein L24
					BGL51_RS09400	50S ribosomal protein L14
					BGL51_RS09405	30S ribosomal protein S17
					BGL51_RS09410	50S ribosomal protein L29
					BGL51_RS09415	50S ribosomal protein L16
					BGL51_RS09420	30S ribosomal protein S3
					BGL51_RS09425	50S ribosomal protein L22
					BGL51_RS09430	30S ribosomal protein S19
					BGL51_RS09435	50S ribosomal protein L2
					BGL51_RS09440	50S ribosomal protein L23
					BGL51_RS09445	50S ribosomal protein L4
					BGL51_RS09450	50S ribosomal protein L3
					BGL51_RS09455	30S ribosomal protein S10
					BGL51_RS09460	Holliday junction branch migration DNA helicase RuvB
					BGL51_RS09465	acyltransferase
					BGL51_RS09470	membrane protein
					BGL51_RS09475	low molecular weight phosphotyrosine protein phosphatase
					BGL51_RS09480	adenylosuccinate synthetase
					BGL51_RS09485	DNA topology modulation protein FlaR
					BGL51_RS09490	MFS transporter
					BGL51_RS09495	MccB-like protein
					BGL51_RS09500	pseudo
					BGL51_RS09505	helix-turn-helix domain-containing protein
					BGL51_RS09510	hypothetical protein
					BGL51_RS09515	ABC transporter ATP-binding protein
					BGL51_RS09520	MutR family transcriptional regulator
					BGL51_RS09525	radical SAM protein
					BGL51_RS09530	ABC transporter ATP-binding protein
					BGL51_RS09535	XRE family transcriptional regulator
					BGL51_RS09540	pseudo
					BGL51_RS09545	DUF3173 domain-containing protein
					BGL51_RS09550	pseudo

Supplementary Table S12B. Potential donors of genomic islands (Gis) identified in the 23 *S. thermophilus* strains by IslandViewer 4*****

Strain	GI	size	start	end	GC %	best blastn hit (excluding same species)	Query Cover (%)	E-value	Identity (%)	Note
ASCC 1275	10	5,605	1,391,298	1,396,903	26.4	*Streptococcus mitis*	86	0	97	this is a common GI among strains presented in column A
DGCC 7710	8	5,607	1,204,419	1,210,026	26.4	*Streptococcus mitis*	86	0	97	
GABA	9	5,607	1,209,343	1,214,950	26.4	*Streptococcus mitis*	86	0	97	
JIM 8232	8	5,607	1,285,617	1,291,224	26.5	*Streptococcus mitis*	86	0	97	
KLDS 3.1003	10	5,607	1,832,431	1,838,038	26.5	*Streptococcus mitis*	86	0	97	
KLDS SM	4	5,607	516,264	521,871	26.4	*Streptococcus mitis*	86	0	97	
LMD-9	6	5,607	1,203,277	1,208,884	26.4	*Streptococcus mitis*	86	0	97	
MN-BM-A02	9	5,607	1,201,043	1,206,650	26.4	*Streptococcus mitis*	86	0	97	
ND07	10	4,400	1,457,288	1,461,688	26.1	*Streptococcus mitis*	83	0	97	
SMQ-301	6	5,607	1,215,056	1,220,663	26.4	*Streptococcus mitis*	86	0	97	
ST3	4	5,607	1,224,942	1,230,549	26.4	*Streptococcus mitis*	86	0	97	
GABA	8	4,426	1,048,971	1,053,397	39.0	*Streptococcus equinus*	99	0	94	this is a common GI among strains presented in column A
LMD-9	5	4,426	1,043,027	1,047,453	38.9	*Streptococcus equinus*	99	0	94	
SMQ-301	5	4,426	1,052,393	1,056,819	38.9	*Streptococcus equinus*	99	0	94	
EPS	5	4,333	951,757	956,090	40.3	*Streptococcus equi* subsp. *zooepidemicus*	87	0	95	this is a common GI among strains presented in column A
S9	2	4,109	126,095	130,204	40.3	*Streptococcus equi* subsp. *zooepidemicus*	87	0	95	
APC151	8	11,868	1,150,322	1,162,190	32.2	*Streptococcus dysgalactiae* subsp. *equisimilis*	85	0	93	this is a common GI among strains presented in column A
ASCC 1275	3	13,289	289,525	302,814	33.0	*Streptococcus dysgalactiae* subsp. *equisimilis*	77	0	93	
CNRZ1066	1	14,523	99,677	114,200	32.5	*Streptococcus dysgalactiae* subsp. *equisimilis*	77	0	92	
CS8	6	19,736	792,107	811,843	34.0	*Streptococcus dysgalactiae* subsp. *equisimilis*	72	0	92	

DGCC 7710	1	12,728	101,144	113,872	32.3	*Streptococcus dysgalactiae* subsp. *equisimilis*	80	0	93	
EPS	1	9,190	120,757	129,947	30.1	*Streptococcus dysgalactiae* subsp. *equisimilis*	86	0	93	
GABA	1	11,221	101,513	112,734	32.1	*Streptococcus dysgalactiae* subsp. *equisimilis*	90	0	93	
KLDS SM	10	13,289	1,262,534	1,275,823	32.9	*Streptococcus dysgalactiae* subsp. *equisimilis*	77	0	93	
LMG 18311	1	13,260	100,771	114,031	31.8	*Streptococcus dysgalactiae* subsp. *equisimilis*	76	0	92	
MN-BM-A01	10	7,918	1,767,123	1,775,041	30.1	*Streptococcus dysgalactiae* subsp. *equisimilis*	99	0	94	
MN-BM-A02	2	13,289	99,521	112,810	32.9	*Streptococcus dysgalactiae* subsp. *equisimilis*	77	0	93	
MN-ZLW-002	2	7,919	105,020	112,939	30.1	*Streptococcus dysgalactiae* subsp. *equisimilis*	99	0	94	
ND03	1	11,869	106,486	118,355	32.1	*Streptococcus dysgalactiae* subsp. *equisimilis*	85	0	93	
ND07	4	15,795	699,018	714,813	33.4	*Streptococcus dysgalactiae* subsp. *equisimilis*	75	0	93	
S9	6	19,736	813,720	833,456	34.0	*Streptococcus dysgalactiae* subsp. *equisimilis*	72	0	92	
LMG 18311	8	13,572	1,585,316	1,598,888	38.1	*Streptococcus equinus*	86	0	95	this is a common GI among strains presented in column A
MN-BM-A01	6	10,188	1,443,892	1,454,080	39.3	*Streptococcus equinus*	94	0	95	
MN-ZLW-002	8	10,188	1,630,277	1,640,465	39.3	*Streptococcus equinus*	94	0	95	

DGCC 7710	11	4,070	1,738,747	1,742,817	38.5	*Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* plasmid pLd7/*Lactococcus lactis* subsp. *lactis* plasmid p229C	100	0	98	this is a common GI among strains presented in column A
KLDS SM	6	4,070	1,050,630	1,054,700	38.4	*Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* plasmid pLd7/*Lactococcus lactis* subsp. *lactis* plasmid p229C	100	0	98	
MN-BM-A02	11	4,070	1,735,404	1,739,474	38.4	*Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* plasmid pLd7/*Lactococcus lactis* subsp. *lactis* plasmid p229C	100	0	98	
ND07	8	4,070	924,460	928,530	38.4	*Lactococcus lactis* subsp. *lactis* bv. *diacetylactis* plasmid pLd7/*Lactococcus lactis* subsp. *lactis* plasmid p229C	100	0	98	
MN-BM-A01	4	4,411	796,536	800,947	42.9	*Lactobacillus casei* plasmid pBD-II / plasmid pLC2W	99	0	98	this is a common GI among strains presented in column A
MN-ZLW-002	6	4,411	982,889	987,300	42.9	*Lactobacillus casei* plasmid pBD-II / plasmid pLC2W	99	0	98	
MN-BM-A01	9	26,007	1,678,515	1,704,522	42.0	*Streptococcus* sp.	96	0	94	this is a common GI among strains presented in column A
MN-ZLW-002	1	26,008	16,411	42,419	42.0	*Streptococcus* sp.	96	0	94	
ND03	10	7,580	1,761,323	1,768,903	34.8	*Streptococcus parasanguinis*	85	0	94	
LMD-9	9	28,974	1,710,761	1,739,735	40.9	*Streptococcus equinus*	99	0	92	
NCTC12958	16	6,025	1,720,856	1,726,881	41.0	*Leuconostoc gelidum* subsp. *gasicomitatum* plasmid 1	71	0	95	

KLDS 3.1003	5	17,978	1,209,205	1,227,183	39.1	*Streptococcus suis*	89	0	97	
NCTC12958	3	10,508	185,061	195,569	37.1	*Streptococcus pneumoniae*	93	0	91	
ACA-DC 2	8	4,150	1,249,165	1,253,315	32.0	*Streptococcus equinus*	100	0	94	unique
B59671	3	4,173	560,018	564,191	40.6	*Streptococcus* sp.	99	0	92	unique
GABA	3	14,916	509,953	524,869	45.2	*Streptococcus macedonicus*	100	0	99	unique
GABA	6	4,668	895,089	899,757	33.6	*Streptococcus infantarius* subsp. *infantarius*	73	0	99	unique
JIM 8232	7	8,390	1,193,475	1,201,865	41.6	*Streptococcus equinus*	98	0	92	unique
JIM 8232	10	6,193	1,494,953	1,501,146	33.3	*Streptococcus equinus*	100	0	99	unique
NCTC12958	2	6,696	113,975	120,671	30.7	*Streptococcus gordonii*	100	0	99	unique
NCTC12958	4	5,047	208,223	213,270	38.9	*Streptococcus suis*	99	0	96	unique
NCTC12958	6	4,635	598,724	603,359	42.9	*Streptococcus macedonicus*	100	0	95	unique
NCTC12958	7	5,906	603,419	609,325	41.7	*Streptococcus agalactiae*	89	0	95	unique
NCTC12958	9	38,494	726,278	764,772	40.5	*Streptococcus* phage 20617	100	0	99	unique
NCTC12958	14	8,397	1,429,880	1,438,277	37.3	*Streptococcus* sp.	85	0	93	unique
NCTC12958	19	23,933	1,840,027	1,863,960	38.9	*Streptococcus suis*	94	0	95	unique
NCTC12958	21	5,730	1,985,007	1,990,737	32.6	*Streptococcus gallolyticus*	99	0	99	unique

***** Potential donors have been identified after blastn analysis (only results with > 70% query coverage and > 90% identity are presented)

Supplementary Table S13. Genes involved in protocooperation with *L. bulgaricus* identified in the 23 *S. thermophilus* strains. Genes highlighted in red correspond to putative pseudogenes

product	gene	ACA-DC 2	APC151	ASCC 1275	B59671	CNRZ1066
Pyruvate formate lyase	*pflB*	STACADC2_RS07510	B1761_RS03780	T303_RS08960	CG712_RS04300	STR_RS07795
Pyruvate formate lyase activating enzyme	*pflA*	STACADC2_RS07420	B1761_RS03690	T303_RS08875	CG712_RS04210	STR_RS07710
DNA starvation/stationary phase protection protein	*dpr*	STACADC2_RS03600	B1761_RS09365	T303_RS04775	CG712_RS00215	STR_RS03680
transcriptional repressor	*fur*	STACADC2_RS03605	B1761_RS09370	T303_RS04780	CG712_RS00220	STR_RS03685
Ferrichrome ABC transporter, substrate-binding protein	*fatB*	STACADC2_RS04830	-	T303_RS06060	CG712_RS01395	STR_RS04980
Ferrichrome ABC transporter, ATP-binding protein	*fatA*	STACADC2_RS04835	-	T303_RS06065	CG712_RS01400	STR_RS04985
Ferrichrome ABC transporter, permease protein	*fatC*	STACADC2_RS04840	-	T303_RS06070	CG712_RS01405	STR_RS04990
Ferrichrome ABC transporter, permease protein	*fatD*	STACADC2_RS04845	-	T303_RS06075	CG712_RS01410	STR_RS04995
Glutathione biosynthesis bifunctional protein gshB	*gshF*	STACADC2_RS06365	B1761_RS02595	T303_RS07800	CG712_RS03155	STR_RS06640

product	gene	CS8	DGCC 7710	EPS	GABA	JIM 8232
Pyruvate formate lyase	*pflB*	BAY21_RS06575	CW339_RS07980	CR922_RS01555	CR921_RS08085	STH8232_RS08225
Pyruvate formate lyase activating enzyme	*pflA*	BAY21_RS06660	CW339_RS07890	CR922_RS01645	CR921_RS07995	STH8232_RS08140
DNA starvation/stationary phase protection protein	*dpr*	BAY21_RS01240	CW339_RS03730	CR922_RS05755	CR921_RS03735	STH8232_RS04025
transcriptional repressor	*fur*	BAY21_RS01235	CW339_RS03735	CR922_RS05750	CR921_RS03740	STH8232_RS04030
Ferrichrome ABC transporter, substrate-binding protein	*fatB*	BAY21_RS09380	CW339_RS05035	CR922_RS04440	CR921_RS05070	-
Ferrichrome ABC transporter, ATP-binding protein	*fatA*	BAY21_RS09375	CW339_RS05040	CR922_RS04435	CR921_RS05075	-
Ferrichrome ABC transporter, permease protein	*fatC*	BAY21_RS09370	CW339_RS05045	CR922_RS04430	CR921_RS05080	-
Ferrichrome ABC transporter, permease protein	*fatD*	BAY21_RS09365	CW339_RS05050	CR922_RS04425	CR921_RS05085	-
Glutathione biosynthesis bifunctional protein gshB	*gshF*	BAY21_RS07720	CW339_RS06815	CR922_RS02750	CR921_RS06875	STH8232_RS07060

product	gene	KLDS 3.1003	KLDS SM	LMD-9	LMG 18311	MN-BM-A01
Pyruvate formate lyase	*pflB*	BEN15_RS01340	A9497_RS04170	STER_RS07960	STU_RS17145	AMD33_RS06860
Pyruvate formate lyase activating enzyme	*pflA*	BEN15_RS01250	A9497_RS04085	STER_RS07875	STU_RS17060	AMD33_RS06775
DNA starvation/stationary phase protection protein	*dpr*	BEN15_RS06955	A9497_RS00040	STER_RS03760	STU_RS13100	AMD33_RS02580
transcriptional repressor	*fur*	BEN15_RS06960	A9497_RS00045	STER_RS03765	STU_RS13105	AMD33_RS02585
Ferrichrome ABC transporter, substrate-binding protein	*fatB*	BEN15_RS08240	A9497_RS01295	STER_RS05060	STU_RS14335	-
Ferrichrome ABC transporter, ATP-binding protein	*fatA*	BEN15_RS08245	A9497_RS01300	STER_RS05065	STU_RS14340	-

Ferrichrome ABC transporter, permease protein	*fatC*	BEN15_RS08250	A9497_RS01305	STER_RS05070	STU_RS14345	-
Ferrichrome ABC transporter, permease protein	*fatD*	BEN15_RS08255	A9497_RS01310	STER_RS05075	STU_RS14350	-
Glutathione biosynthesis bifunctional protein gshB	*gshF*	BEN15_RS00060	A9497_RS03015	STER_RS06790	STU_RS15985	AMD33_RS05710

product	**gene**	**MN-BM-A02**	**MN-ZLW-002**	**NCTC12958**	**ND03**	**ND07**
Pyruvate formate lyase	*pflB*	MNA02_RS07865	Y1U_RS07880	DQL34_RS09110	STND_RS07870	BBD27_RS06175
Pyruvate formate lyase activating enzyme	*pflA*	MNA02_RS07780	Y1U_RS07795	DQL34_RS09020	STND_RS07785	BBD27_RS06260
DNA starvation/stationary phase protection protein	*dpr*	MNA02_RS03690	Y1U_RS03620	DQL34_RS04645	STND_RS03670	BBD27_RS00635
transcriptional repressor	*fur*	MNA02_RS03695	Y1U_RS03625	DQL34_RS04660	STND_RS03675	BBD27_RS00630
Ferrichrome ABC transporter, substrate-binding protein	*fatB*	MNA02_RS04970	-	-	-	BBD27_RS09055
Ferrichrome ABC transporter, ATP-binding protein	*fatA*	MNA02_RS04975	-	-	-	BBD27_RS09050
Ferrichrome ABC transporter, permease protein	*fatC*	MNA02_RS04980	-	-	-	BBD27_RS09045
Ferrichrome ABC transporter, permease protein	*fatD*	MNA02_RS04985	-	-	-	BBD27_RS09040
Glutathione biosynthesis bifunctional protein gshB	*gshF*	MNA02_RS06705	Y1U_RS06725	DQL34_RS07820	STND_RS06685	BBD27_RS07330

product	**gene**	**S9**	**SMQ-301**	**ST3**
Pyruvate formate lyase	*pflB*	AVT04_RS06735	SMQ301_RS07945	BGL51_RS08180
Pyruvate formate lyase activating enzyme	*pflA*	AVT04_RS06820	SMQ301_RS07860	BGL51_RS08085
DNA starvation/stationary phase protection protein	*dpr*	AVT04_RS01315	SMQ301_RS03755	BGL51_RS03760
transcriptional repressor	*fur*	AVT04_RS01310	SMQ301_RS03760	BGL51_RS03765
Ferrichrome ABC transporter, substrate-binding protein	*fatB*	AVT04_RS00065	SMQ301_RS05040	-
Ferrichrome ABC transporter, ATP-binding protein	*fatA*	AVT04_RS00060	SMQ301_RS05045	-
Ferrichrome ABC transporter, permease protein	*fatC*	AVT04_RS00055	SMQ301_RS05050	-
Ferrichrome ABC transporter, permease protein	*fatD*	AVT04_RS00050	SMQ301_RS05055	-
Glutathione biosynthesis bifunctional protein gshB	*gshF*	AVT04_RS07885	SMQ301_RS06775	BGL51_RS06985