Università degli Studi di Padova

Padua Research Archive - Institutional Repository

Genome comparison and physiological characterization of eight *Streptococcus thermophilus* strains isolated from Italian dairy products

Original Citation:

Availability:

This version is available at: 11577/3207678 since: 2017-12-11T09:41:52Z

Publisher:

elsevier

Published version:

DOI: 10.1016/j.fm.2016.11.002

Terms of use:

Open Access

This article is made available under terms and conditions applicable to Open Access Guidelines, as described at <http://www.unipd.it/download/file/fid/55401> (Italian only)

(Article begins on next page)

Genome comparison and physiological characterization of eight *Streptococcus thermophilus* strains isolated from Italian dairy products

Veronica Vendramin ^a, Laura Treu ^{ab}, Stefano Campanaro ^c, Angiolella Lombardi ^a, Viviana Corich ^{a*}, Alessio Giacomini ^a

^a Department of Agronomy, Food, Natural resources, Animal and Environment (DAFNAE), viale dell'Università, 16, 35020 Legnaro (Padova), Italy

^b Department of Environmental Engineering, Technical University of Denmark, Kongens Lyngby, Denmark

^c Department of Biology, University of Padova, via Ugo Bassi 58/b, 35121 Padova, Italy

* Corresponding author. Present address: Department of Agronomy, Food, Natural resources, Animal and Environment (DAFNAE), viale dell'Università, 16, 35020 Legnaro (Padova), Italy. E-mail address: viviana.corich@unipd.it (V. Corich).

ABSTRACT

Eight *Streptococcus thermophilus* strains of dairy origin isolated in Italy were chosen to investigate autochthonous bacterial diversity in this important technological species. In the present study a comparative analysis of all the 17 *S. thermophilus* genomes publicly available was performed to identify the core and the variable genes, which vary among strains from 196 to 265. Additionally, correlation between the isolation site and the genetic distance was investigated at genomic level. Results highlight that the phylogenetic reconstruction differs from the geographical strain distribution. Moreover, strain M17PTZA496 has a genome of 2.15 Mbp, notably larger than that of the others, determined by lateral gene transfer (including phage-mediated incorporation) and duplication events. Important technological characters, such as growth kinetics, bacteriocin production, acidification kinetics and surface adhesion capability were studied in all the Italian strains. Results indicate a wide range of variability in adhesion properties that significantly clustered strains into four groups. Genomic differences among strains in relation to these characters were identified but a clear correlation between genotype and phenotype was not always found since most of the genomic modifications arise from single nucleotide polymorphisms. This research represents a step forward in the identification of strains-specific functions in *Streptococcus thermophilus* and it

34 has also the potential to provide valuable information to predict strain specific behaviors in industrial
35 processes.

36
37
38

39 **KEYWORDS**

40
41
42

Dairy products, *Streptococcus thermophilus*, bacterial biodiversity, genome comparison

43 **Highlights**

- 44 • Genome comparison of all *S. thermophilus* strains publicly available was performed
- 45 • Weak correlation between genome evolution and geographical origin was found
- 46 • Extracellular polymeric substances and bacteriocin sequences are present but phenotypes
47 were not detected
- 48 • Acidification kinetics were highly variable among strains isolated in Italy
- 49 • Technological implications of *S. thermophilus* genomic diversity was evaluated

50

51 **1. Introduction**

52 *Streptococcus thermophilus* is a thermophilic Lactic Acid Bacterium (LAB) of major importance in
53 the dairy industry. Due to its ability to rapidly ferment lactose, it is widely used as starter to obtain
54 fermented milk products, contributing to milk acidification and organoleptic properties enrichment
55 (Giraffa et al., 2001). The long history of safe use in food production allowed *S. thermophilus* to
56 obtain the status of Generally Recognized as Safe (GRAS) and of Qualified Presumption of Safety
57 (QPS). At present, it is considered the second most important species of industrial LAB after
58 *Lactococcus lactis*. It was estimated that over 10^{21} live cells are ingested annually leading the species
59 to achieve a market value of 40 billion US\$, approximately (Iyer et al., 2010). Similarly to other dairy
60 microbes, *S. thermophilus* natural biodiversity decreases with its overuse of industrial starters, hence
61 isolation and characterization of new strains becomes of great importance, since it may lead to the
62 discovery of novel and desirable characteristics, which can fulfil industrial demands (Erkus, Okuklu,
63 Yenidunya, & Harsa, 2014). Strains analyzed in the present study are used as natural starters for
64 Protected Designation Origin (PDO) Italian cheeses, i.e. Fontina, Grana Padano and Mozzarella.
65 These products, obtained from traditional back-slopping procedures, allow the maintenance of the
66 microbiota present in the environment where they are produced. Considering the number of Italian
67 cheese factories involved in dairy productions (ISTAT, 2014), the Italian dairy microbiota can be
68 considered a potential important source of new *S. thermophilus* strains. Nonetheless, until now such
69 biodiversity has been explored only partially both from the genetic and phenotypic point of view
70 (Andrighetto et al., 2002, Morandi and Brasca, 2012). Thanks to the next generation sequencing
71 technology, whole *S. thermophilus* genome sequences are publicly available. Such information
72 allowed to study more in depth the genetic structure of many metabolic activities of the species, such
73 as amino acid metabolism (Hols et al., 2005), arrangement of the proteolytic system (Hols et al.,
74 2005), resistance to bacteriophage (Li et al., 2016), biosynthesis of folate (Iyer et al., 2010),
75 metabolism of urea (Mora et al., 2004) and biofilm formation (Couvigny et al., 2015).
76 More generally, *S. thermophilus* genomes have been so far analyzed and compared with other related
77 species, specifically with pathogenic streptococci (Hols et al., 2005). Within this framework,
78 Rasmussen et al. (2008) used microarrays to perform a comparative genomic analysis of different *S.*
79 *thermophilus* strains to demonstrate the presence of variable subsets of genes responsible for
80 ecological and technological differences. One of the most interesting technological properties
81 reported in comparative studies is related to the ability to synthesize extracellular polymeric
82 substances (EPS) (Flemming and Wingender, 2010, Mora et al., 2002). Recently, the beneficial
83 effects of EPS in fermented milk have been recognized and linked to their role as thickeners and

84 stabilizers of the product, together with healthy effects, such as their activity on human blood pressure
85 and gastrointestinal tract health (Caggianiello et al., 2016). To date, 28 different EPS gene clusters
86 are known in *S. thermophilus* (Iyer et al., 2010, Vuyst et al., 2011, Wu et al., 2014).
87 This species has been tested as bio-preservative to control growth of pathogenic and spoilage bacteria
88 in dairy products by production of bacteriocins (Kongo, 2013). These molecules, produced by food
89 grade lactic acid bacteria, are classified into two classes based on their modification status. Known
90 *Streptococcus* bacteriocins belong to class I and class IIb (Egan et al., 2016) and the identification of
91 genes encoding for bacteriocins is a quite difficult task due to their small size and high variability in
92 sequence composition (Willey and van der Donk, 2007).
93 The present study used the results obtained thanks to the most innovative Next Generation
94 Sequencing approach (Treu et al., 2014a, Treu et al., 2014b, Treu et al. 2014c) to provide more details
95 on the genetic organization of *S.thermophilus* at whole genome level. The overall biodiversity of
96 Italian *S. thermophilus* was studied by comparing eight isolates, coming from PDO cheese
97 productions. It is known that subsets of features specifically characterizing different strains are
98 extremely important when they are forced to face environmental changes (Hols et al., 2005). The
99 geographical effect on biodiversity was examined by comparing the genome sequences of all strains
100 available in the literature. Furthermore, technological characters related to food production were
101 investigated in the Italian strains and linked to genomic data.
102

103 2. Material and methods

104 2.1. Strain used

105 Considering all surveys regarding *S. thermophilus* species, seventeen genomes publicly available in
106 the NCBI database in August 2015 (Table 1) were used. Other species used as outgroups or references
107 for specific analysis are described in the correspondent paragraphs. In the present study gene finding
108 and annotation for all the strains were newly performed using RAST (Rapid Annotation using
109 Subsystem Technology) service which orders annotated genes into subsystems, subcategories and
110 categories, following the SEED structure (Aziz et al., 2008). Gene function is assigned by sequence
111 attribution to protein families (FIGfams). Eight strains isolated from Italian dairy products were used
112 for phenotypic tests (Table 1). Unless otherwise indicated, *S. thermophilus* strains were grown
113 overnight at 37°C in M17 broth (Oxoid, UK). For long-term maintenance, grown cultures were stored
114 at -80 °C in 40% (v/v) glycerol and 5% skim milk (Sigma-Aldrich, Italy).

115 2.2. Phylogenetic and genomic analyses

116 Genomic data of 17 *S. thermophilus* genomes along with *Streptococcus macedonicus* 33MO,
117 *Streptococcus pneumoniae* NT_110_58 and *Streptococcus salivarius* JIM8777 were used to estimate
118 phylogenetic relationships by combining two methods. The first phylogenetic tree was obtained using
119 PHYLIP package (Tuimala, 2005) with neighbour-joining method. This method utilized single
120 nucleotide polymorphisms (SNPs) of the whole genome alignment computed using Mauve software
121 (Darling et al., 2004) with a procedure previously described (Treu et al., 2014). In the second case,
122 the phylogenetic tree was built using PhyloPhlAn software (Segata et al., 2013) which determines
123 microbial phylogeny on the basis of 400 conserved proteins alignment. Phage proteins were
124 recognized by RAST gene functional attribution and their organization was manually explored
125 localizing sequences coding for phage proteins. Laterally transferred regions were identified using
126 Alien Hunter software (Vernikos and Parkhill 2006). From the output, sequences putative functions
127 were identified by blastp alignment. Gene duplication analysis was performed according to
128 Campanaro et al. (2014) by clustering total strains proteins using CD-HIT software (Li and Godzik,
129 2006). Two different analyses were performed using 90% and 99% identity of sequence and minimal
130 length similarity of “0.5”.

131 2.3. Gene content evaluation

132 For each *S. thermophilus* strains, annotated gene were attributed to subsystem and a features list was
133 created based on subsystem gene abundance. Feature lists were used to elaborate hierarchical

134 clustering (HCL) using MeV (MultiExperiment Viewer) software (Saeed et al., 2003). Strains
135 functional relationship was computed using the “linkage method” process for determining cluster-to-
136 cluster distances and the “Euclidean distance” for distance calculation. A comparison on subsystem
137 gene abundance was conducted analyzing the resulting heatmap and the most interesting subsystems
138 were manually investigated in detail. For the specific subsystem ‘Iron acquisition and metabolism’,
139 the following strains of *Streptococcus pyogenes* were used: M1GAS, MGAS10270, MGAS10394,
140 MGAS1075, MGAS2096, MGAS315, MGAS5005, MGAS6180, MGAS8232, MGAS9429, SSI-1
141 and str. Manfredo. *Lactobacillus fabifermentans* T30PCM01 genome was also included in the
142 analysis (Treu et al., 2014d). On the basis of features lists, non-redundant common and non-common
143 strain features were identified using R software, custom script (R Development Core Team, 2008).
144 Strain specific features were assigned to the SEED categories using RAST.

145 2.4. Identification of sequences related to technological properties

146 *S. thermophilus* proteolytic activity was studied by verifying sequence presence of species specific
147 main components, namely the cell-envelope protease, Ptrs, and the protein responsible for its
148 anchoring to bacterial membrane, Sortase A (SrtA). Sequences of *S. thermophilus* MN-ZLW-002
149 (YP_006340201.1 and YP_006340309 for Ptrs and Srt A respectively) were used to perform blastp
150 search using strains genome as reference sequence.

151 Exopolysaccharides-related genes were analyzed considering subsystems completeness and sequence
152 similarities. Genes assigned to “EPS” and “CPS” were identified for each strain, their number of
153 copies and the organization into operons were recorded. Sequences were clustered using CD-HIT at
154 50%, 80%, 90% and 100% of identity in order to understand their degree of similarity. Promoter
155 regions of the main EPS operon of each strain were compared against a reference strain, ND03,
156 previously studied for its abilities in EPS production (Sun et al., 2011). Bacteriocins are important
157 strain specific compounds and BAGEL3 software (van Heel et al., 2013) was chosen to determine
158 microorganism potentiality to produce novel compounds. Putative biosynthetic gene clusters were
159 identified in the genome sequence (Egan et al., 2016).

160 2.5. Strain growth curve

161 For growth curve determination, a loopful of cells from frozen stocks was inoculated in 10 ml of M17
162 broth pre-warmed at 37°C and incubated at 37°C overnight. The cultures were used to inoculate 200
163 ml of fresh M17 broth to a concentration of 10^5 cells/ml and bacterial growth was monitored by plate
164 counts. To estimate growth parameters, 3 different mathematical models were used, namely
165 Gompertz (Zwietering et al, 1990), Baranyi (Baranyi and Roberts, 1994) and Huang (Huang, 2008).

166 Models goodness-of-fit were evaluated using four indexes, namely bias factor (BF), accuracy factor
167 (AC), sum of squares error (ESS) and AIC index according to Huang 2010 in order to establish the
168 most suitable model for growth parameters prediction. Data were analyzed using XLSTAT (version
169 2011, Addinsoft, USA).

170 *2.6. Acidification kinetics and protease activity determination*

171 A loopful of bacteria from freshly grown M17 plates was used to inoculate 10 ml of 10% (w/v) skim
172 milk (Oxoid, IT), previously sterilized by autoclaving 10 min at 110°C. Inoculated samples were kept
173 in a water bath at 42°C until milk coagulation. Only strains able to coagulate milk within 16 hours
174 were further tested, as follows.

175 Flasks containing 250ml of sterilized 10% (w/v) skim milk (Oxoid, IT) were inoculated with 2% (v/v)
176 of cultures obtained as describe above. Flasks were incubated in a water bath at 37°C and the pH
177 recorded continuously for 24h by a pH electrode (Micros, Siap+Micros, Treviso, Italy) immersed in
178 the medium and connected to a software system for data acquisition. According to Dandoy and
179 colleagues (2011) results were expressed as maximum acidification rate (V_m), defined as the maximum
180 slope of the pH curve (dpH/dt) and time required to reach two pH values, pH 5.2 and 4.6.

181 The PrtS proteinase phenotype was phenotypically determined on bacterial colonies grown on semi
182 skimmed milk agar plate according to Morris et al. (2012). For each strain, 5 μ l of a routinely grown
183 culture were dropped on lactose-free skim milk plates and incubated at 37°C for 24h and 48h. Strains
184 were considered to express protease activity when a transparent halo appears around the culture drops.
185 Each experiment was repeated at least 3 times and statistical analyses were performed using R
186 software.

187 *2.7. Exopolysaccharides production and antimicrobial activity*

188 Strains were tested for EPS production using a colorimetric assay. Strains were routinely grown in
189 liquid M17 broth and used to inoculate a fresh M17 broth (1% v/v) dispensed into microtiter plate
190 wells (200 μ l per well). Cultures were incubated at 37°C and the increase in absorbance (OD₆₀₀) was
191 monitored every 30 min, after gently shaking for 30 sec. After 24h, biofilm formation was quantified
192 as described by Maragkoudakis et al. (2013). OD value recorded for empty wells plus three standard
193 deviations were used as control. Results were evaluated by ANOVA tests performed by R. For
194 scanning electron microscope (SEM) analysis, routinely grown cultures were refreshed by adding
195 10% (v/v) of fresh medium into sterile Petri dish containing a glass coverslip and statically incubated
196 at 37 °C for 24 h. Coverslips provide the adhesion surface for bacterial cells, therefore they were
197 recovered after M17 broth gently removal and PBS buffer washing (NaCl 137 mM, KCl 2.7 mM,

198 Na_2HPO_4 10 mM, KH_2PO_4 2 mM, pH 7.4). Sample fixation, dehydration and assembly were
199 performed as described by Campanaro et al., 2014. Samples were observed and photographed by a
200 Quanta 200 SEM (FEI, Hillsboro, OR, USA).
201 Antimicrobial activity was determined by agar-spot test (Rossi et al., 2013). Details on strains and
202 protocols are reported in the Supplementary information.
203

204 3. Results and discussion

205 3.1. Phylogenetic reconstruction

206 Eight *S. thermophilus* strains isolated from Italian dairy products were sequenced to investigate
207 autochthonous bacterial diversity of this important technological starter species (Giraffa et al., 2001,
208 Treu et al., 2014a, Treu et al., 2014b, Treu et al. 2014c). In the present study a comparative analysis
209 of all *S. thermophilus* genomes presently available in public databases (Table 1) was performed.
210 Although all strains were collected from dairy environments, they derived from different food-making
211 processes and from milk of different mammals (Table 1). Two independent phylogenetic analyses
212 were performed, one based on the conserved proteins and another based on single nucleotide
213 polymorphisms (SNPs) detected comparing the whole genome sequences. . Since M17PTZA496 was
214 found to possess a large strain-specific region which was excluded from the whole genome
215 computation, the second tree was aimed to avoid this bias. The two resulting phylogenetic trees are
216 concordant in defining the relationships among strains (Fig. 1). It is clear that the phylogenetic
217 reconstruction does not follow geographic distribution since strains isolated in the same continent are
218 rarely clustered together. Indeed, strains isolated in Europe are widespread on the phylogenetic tree
219 while American and Asian strains show a higher proximity. Within European strains, some Italian
220 strains are phylogenetically close, namely TH982/TH985 isolated in Campania region and
221 TH1435/TH1436 isolated in Friuli Venezia Giulia region. These two strain couples were collected
222 from the same food matrices, buffalo and goat milk respectively (Table 1). Therefore it cannot be
223 excluded that both factors contributed in determining genome similarity.
224 Finally, both phylogenetic approaches resulted in a concordant topology indicating higher distance
225 of strain M17PTZA496 from the others (Fig. 1). I can be hypothesized that this strain is undergoing
226 an evolutionary process that could eventually resolve into its diversification.

227 3.2. Genetic diversity of strain M17PTZA496

228 Genetic diversity ascribed to strain M17PTZA496 depends from the highest number of SNPs and
229 from the presence of a large strain-specific genomic region. In fact, the 17 genomes considered in the
230 present study have a comparable size (from 1.93 to 1.74 Mbp) with the notable exception of strain
231 M17PTZA496 (2.15 Mbp), which carries almost 0.3 Mbp more genetic information than the average
232 (Table 1). To clarify the origin of M17PTZA496 differentiation, lateral gene transfer (including
233 phage-mediated incorporation) and gene duplication event were investigated.
234 Phages are widespread in dairy environment and therefore phage resistance systems are considered a
235 technological character of major interest (Goh et al., 2011). It is known that several genes belonging

236 to CRISPR/cas system increase their expression during response to bacteriophage attack (Wu et al.,
 237 2014). Interestingly, when compared with the other strains, M17PTZA496 shows the lower number
 238 of CRISPR/cas genes (Table 1). Moreover, M17PTZA496 contains a higher number of unknown
 239 proteins, or other related to “phages/prophages” category proteins compared to the other strains: 81
 240 proteins in M17PTZA496 and from 5 to 15 in the others. Most of these proteins are positioned in a
 241 phages-rich protein region (PH) spanning approximately 42 Kbp in *contig23* (from 1266 to 43435
 242 bp; Fig. 2), related to phage functioning (i.e. tail proteins, replication, packaging machinery, tail fiber
 243 proteins, capsid proteins and lysis). From these results it’s possible to hypothesize that the reduction
 244 in CRISPR/cas genes content of M17PTZA496 strain has increased its susceptibility to phage
 245 infection resulting into acquisition of new genetic material.

246 Analysis of laterally transferred regions (LTR) performed by Alien Hunter software led to the
 247 identification of four LTRs, two of which consisting in large genomic islands (Fig. 2), named ISL1
 248 and ISL2, with a size of 37.5 and 77.5 Kbp respectively (in *contig11* from 192500 to 230000 bp and
 249 in *contig61* from 1462450 to 1540000 bp). Results of sequences similarity search demonstrated that
 250 the acquired genes are not closely related but span a wide range of different functions, including
 251 transport and stress response (Supplementary Table S1). While ISL1 encodes features clearly
 252 recognized as part of the normal genetic pool of the *Streptococcus* genus, on the contrary ISL2
 253 includes genes having similarity with different species. Moreover, similarity values obtained are
 254 lower in ISL2, which could indicate a higher decay rate or a stronger selective pressure on the second
 255 region (i.e. on average 99% of identity with E-values of 7.7E-19 and 70% of identity with E-values
 256 4.8E-6 respectively). Based on these results it is possible to hypothesize that ISL2 was included in
 257 M17PTZA496 genome prior to ISL1. The two smaller regions, “Island3” and “Island4”, have a size
 258 of 15 Kbp and 7.5 Kbp respectively and large part of the proteins present in these regions have
 259 unknown functions.

260 Together with LTR and gene loss, gene duplication and regulatory divergence of paralogues are
 261 fundamental in determining bacterial functional and technological properties (Snel et al., 2002). A
 262 specific analysis was performed on duplicated genes identified on the Italian isolates, in order to
 263 investigate M17PTZA496 paralogues content. Present findings revealed 60 clusters of genes with
 264 similarity >90% (6 with similarity >99%), mainly assigned to mobile elements and related proteins
 265 (Supplementary Table S2). Considering the paralogues clusters with >90% similarity, three groups
 266 were arbitrarily defined according to the number of strains having the same paralogues. In class **(I)**
 267 several strains possess duplicated genes in cluster, in class **(II)** only strains from Fontina have
 268 duplicated genes in cluster while the other have only one copy of the sequence, and in class **(III)**
 269 duplicated genes are present in strain M17PTZA496 while in the other strains there is only a single

copy of the gene. This classification revealed that 30% of paralogues were found in at least three strains. Interestingly, a small amount of unknown genes (5%) is present in both strains isolated from the same environment (M17PTZA496 and MTH17CL396). More than half of the paralogues clusters (63%) are present only in M17PTZA496 and most of them are ribosomal proteins while others have specific functions (e.g. UDP-glucose 4-epimerase) (Supplementary Table S2). The remaining 8% of paralogues is not included in the above classification because they are duplicated mobile elements exclusively present in strain M17PTZA496 and absent in the other strains. Finally, 18 out of 60 paralogues clusters belonging to class **(III)** are composed by genes located in a specific genomic region (DR) in *contig69* (from 82019 to 93062 bp), which probably underwent a single duplication event producing the second copy located in *contig71* (from 7519 to 18640 bp; Fig. 2).

3.3. Functional variability

Phenotypic differences among strains could be determined by variation in the number of genes assigned to specific SEED functional categories.. Hierarchical clustering was performed on the profiles reporting the number of genes for each functional category in order to evaluate similarities among 17 strains. This analysis was also useful to identify correlations between gene content and strain geographical isolation sites (Fig. 3). Variations in gene abundance of each functional category were used to build a dendrogram reporting the level of divergence between strains (Fig. 3).

Results obtained from hierarchical clustering are highly different from the phylogenetic reconstruction. In particular, in the “functional clustering” Italian strains formed a compact cluster, while they were found to be distantly related from a phylogenetic point of view. Only strains TH982 and TH985 clustered together with strains distantly isolated, forming a branch separate from the others.

Besides SEED categories describing basic metabolism, such as protein, RNA and DNA metabolism, a subset of functional categories are particularly important to define distinctive characters among *S. thermophilus* strains . These highly variable categories were organized for discussion purposes into four groups: **(a)** “Cofactors, vitamins, prosthetic groups and pigments” **(b)** “Cell wall and capsule” and “Membrane transport”, **(c)** “Stress response” and **(d)** “Nitrogen metabolism” and “Amino acids and derivatives” (Fig. 3).

The first group **(a)** includes the “Cofactors, vitamins, prosthetic groups pigments” category and it is mainly due to variation in “Riboflavin, FMN, FAD” and “Folate and pterines” subcategories, which are involved in the production of the most important cofactors, key targets for the development of new vitamin-enriched products (Russo et al., 2014; Divya and Nampoothiri, 2015).

Three SEED subcategories, “capsular and extracellular polysaccharides”, “cell wall and capsule” and “membrane transporter” are the most relevant in the second group **(b)**. Their importance derives to the role of extracellular polysaccharides (EPS) in determining technological characteristics (see par. 3.6.), such as organoleptic and healthy properties of the fermented end-products (Awad et al., 2005). Interestingly, different strains evidenced a high variability in the number of genes of the third group **(c)**, represented mainly by three subcategories: “osmotic stress”, “oxidative stress” and “detoxification”. In detail, concerning “osmotic stress”, genes related to choline and betaine uptake and biosynthesis were specifically found in 8 strains (Fig. 3). Under standard conditions, Gram-positive bacteria possess turgor pressure higher than Gram-negative and respond to hyper-osmotic condition accumulating protective compounds such as glycine betaine (Sleator and Hill, 2001). The fourth group **(d)** is characterized by genes involved in nitrogen and amino acids metabolism. The main differences are related to “lysine, threonine, methionine and cysteine”, “nitrogen metabolism” and “histidine metabolism”, which ranged from 0 to 14 genes depending on strain. This finding is in contrast with a previous study revealing a strong conservation in the amino acids metabolism among strains (Hols et al., 2005).

3.4. Strain-specific features

A global comparison of the whole gene content among strains of the *S. thermophilus* species allowed the identification of strain specific features, varying in number from 196 to 265, representing 17-21% of the total genes with known function (Fig. 4; Supplementary Table S3). Four functional categories accounted for a large part of strain diversity, namely “amino acids and derivatives”, “carbohydrates”, “DNA metabolism” and “membrane transport” covering almost 50% of the specific genes (11, 12, 13 and 10% respectively). The contribution of these categories to strain variability was in accordance with previous findings obtained by comparative genome hybridization experiments (Rasmussen et al., 2008). Moreover, one of the most interesting subcategories found is involved in stress response (Supplementary Table S3). In detail, only CNRZ1066 and LMG 18311 possess the “acid resistance” subcategory, composed by four genes encoding the glutamate transporter and known to be involved in acid tolerance. These genes were specifically detected in species belonging to the *Streptococcus* genus (Krastel et al., 2010). Other important subcategories describing strain specialization are “oxidative stress”, which includes proteins involved in the protection from reactive oxygen species (ROS) and glutathione homeostasis.

High variability is also affecting the “iron acquisition and metabolism” category, which is known to be connected with oxidative stress response in *Streptococcus* pathogenic species (Tsou et al., 2010). A comparison was performed among *S. thermophilus* strains, 12 *Streptococcus pyogenes* strains

336 available in the RAST database, selected for the presence of several iron related genes (Ge, Sun, &
337 He, 2009), and *Lactobacillus fabifermentans* T30PCM01 that is the *Lactobacillus* strains with the
338 largest genome so far identified (Campanaro et al., 2014). It was reported that adaptation to the dairy
339 environment probably influenced iron requirement of lactic acid bacteria, in fact bacterial growth is
340 unaffected by iron deprivation (Pandey et al., 1994). Indeed, the analyses of *S. pyogenes* strains
341 underlines that this species possesses from 15 to 18 genes ascribed to iron metabolism while *L.*
342 *fabifermentans* has only 5. Strain comparison performed among *S. thermophilus* strains revealed a
343 strong heterogeneity in the number of genes related to iron metabolism which varies between 3 and
344 15. In particular, eight strains showed a strong reduction in iron related gene content, possessing only
345 3 copies of the ferrous iron transporters (*EfeUOB*) and having completely lost the “Heme, hemin
346 uptake and utilization systems in Gram positives” and “Iron acquisition in *Streptococcus*”
347 subcategories.

348 3.5. Acidification kinetics

349 Growth dynamics strictly affect the technological behavior of *S. thermophilus* strains along with their
350 capability to rapidly acidify the milk. Three mathematical models were taken into account to describe
351 bacterial growth kinetics and Gompertz model was selected for all the strains with the exception of
352 MTH17CL396. A model fit assessment criteria, based on the MTH17CL396 growth curve, led to the
353 selection of the Huang model for this strain (Huang, 2010). Growth curves results showed interesting
354 differences among strains (Table 2), considering the three main parameters: lag phase (λ), maximum
355 growth rate ($\mu_{\max}$) and stationary phase concentration ($N_{\max}$). Strain 1F8CT showed the longest lag
356 phase, the higher $\mu_{\max}$ and the lowest stationary phase cell concentration, thus appearing the least
357 performing strain from a technological point of view. On the contrary, strain M17PTZA496 showed
358 the shortest lag phase and the highest stationary phase cell concentration, while TH1477 had the
359 highest $\mu_{\max}$.

360 Acidification capability is important to assure a good outcome of dairy processes and to enhance food
361 safety by hampering the development of undesirable bacteria (Gaden et al., 1992). To evaluate strain
362 performances, fermentation kinetics and acidification capability were monitored in skim milk
363 medium. A preliminary test was performed to assess whether strains were capable to induce
364 coagulation of caseins within 16 h incubation. Only strain 1F8CT failed and was excluded from the
365 following analysis, also considering its bad results in terms of growth kinetics.

366 Results obtained from 24 h acidification kinetics show that all strains reached pH 5.2 (Table 2) and 5
367 out of 7 went below 4.6, which is the value required for caseins coagulation and for the inhibition of
368 pathogen growth. Maximum acidification rate ($V_{\max}$) analysis showed no significant differences in

369 strains acidification profile, nevertheless ANOVA analyses reveal statistical differences in the time
370 required to reach the two pH point considered. Dandoy et al. (2011) reported a significant correlation
371 between acidification kinetics and activity of PtrS, a cell-envelope proteinase (CEP) of *S.*
372 *thermophilus*. In the *Streptococcus* genus, CEPs are generally anchored to the cell wall by sortase A
373 (SrtA) and recently it was reported that *S. thermophilus* proteinase could be released in the medium
374 (Chang et al., 2012). To better understand whether this proteolytic system is present and active in the
375 strains under evaluation, sequence similarity analyses were performed. Results obtained by
376 comparing the reference amino acid sequences of strains MN-ZLW-002 with strains' genotype
377 indicate the presence of SrtA in all strains, with average identity of 97%. On the contrary, PrtS gene
378 sequence was detected only in TH1435, with some minor differences when compared to the reference
379 strain.

380 3.6. Exopolysaccharides production

381 The genomic organization of the genes involved in exopolysaccharides (EPS) production could be
382 summarized as follows: **(1)** regulatory genes (*epsA*, *epsB*), **(2)** genes involved in determining the
383 number of repeated units and their export (*epsC*, *epsD*), **(3)** genes related to the biosynthesis of the
384 repeated units for exopolysaccharide synthesis (*epsE*, *epsF*, *epsG*, *epsH*, and *epsI*), and **(4)** genes
385 active in the polymerization and export (*epsK*, *epsL*, *epsM*) (Iyer et al., 2010)

386 Gene content analysis performed on the EPS-subsystems on the eight *S. thermophilus* strains isolated
387 in Italy showed several differences (Table 3). It is known that the presence of gene sequences coding
388 for the exopolysaccharide biosynthesis transcriptional activator (EpsA), manganese-dependent
389 protein-tyrosine phosphatase (EpsB) and tyrosine-protein kinase transmembrane modulator (EpsC)
390 is not sufficient to determine the “ropy” phenotype (Stingle, Neeser, & Mollet, 1996). Results
391 highlight the presence in all the strains of a “core” subset constituted by 5 genes, namely the previous
392 cited genes together with tyrosine-protein kinase (EpsD) and undecaprenyl-phosphate galactose
393 phosphotransferase (rfbP), indicating their putative role in cell adhesion and aggregation (Sun et al.,
394 2011). Nevertheless, significant variations in gene copy number and sequence similarity were
395 observed in the “core” genes (e.g. TH982 vs TH1435).

396 Additionally, there are other EPS-related genes, generally called glycosyltransferase (*gtf*) (Stingle
397 et al., 1996) and *cps* gene cluster, known to have sequence similarity with *eps* (Bolotin et al., 2004).
398 Results indicated that most of the studied strains possess *gtf* gene sequences (Table 3). Moreover,
399 even though genes belonging to **(3)** and **(4)** were not detected, their homologous *cps* sequences are
400 present in 1F8CT, M17PTZA496 and TH982. As a concluding remark, genetic results evidenced
401 these strains as promising EPS producers, in particular strains TH982 possessing 13 EPS-related

genes. An additional investigation performed on the arrangement of these genes revealed that they appear to be organized in three operons. The main important operon, composed of the “core” genes, was found in all the analyzed strains, while two additional operons encoding the *cps* genes were present only in 1F8CT and in M17PTZA496. A more detailed analysis performed on the promoter region of the “core” operon revealed a 34 bp insertion located between the -35 and -10 regions in strain TH982 (Fig. 5a).

The insertion in the promoter region of strain TH982, absent in the other strains and in the well-known EPS producer strain ND03 (Sun et al., 2011), suggests a possible response for its low production notwithstanding its richness in EPS related genes.

An in-vitro surface adhesion test was performed and highlighted significant differences among strains ($P < 0.001$; Fig. 5b). Statistical analysis organized strains into four groups according to their adhesion capacity: *a*) absent, *b*) low, *c*) medium and *d*) high. Only in MTH17CL396, the best performing strain, results were statistically significant.

Finally, SEM images of MTH17CL396 and TH985 (representatives of high and low adhesion capability, respectively) provided additional information on EPS and allowed to correlate adhesion results with the amount of EPS produced. No clear biofilm formation was visible in SEM images but a variation in cell-to-cell binding capability is clearly evident. In particular, MTH17CL396 displayed the highest robustness in cell anchorage, reflected in a more complex structure and cell chains interaction (Fig. 5c and 5d). Even in the absence of complex extracellular matrix production, a strong cell-to-cell binding could be relevant in producing compact clusters that could help protecting cells during gastrointestinal transit. In this direction, further analyses could lead to discover interesting gastrointestinal survival properties of strain MTH17CL396 due to its highest ability in cell aggregation.

3.7. Bacteriocin production

S. thermophilus strains are normally well represented in the spontaneous microbial consortia of artisanal cheeses and fermented milks. This is in part favored by the production of organic acids and bacteriocins that hamper proliferation of competitor microorganisms (Morandi & Brasca, 2012).

Phenotypical test performed against a selection of marker bacteria (similar to pathogens or food spoilers) and against three lactic acid bacteria did not reveal any inhibitory activity. Conversely, whole genome analysis based on bioinformatics mining by BAGEL3 software (van Heel et al., 2013) identified several lantibiotic-related (LR) genes in all the strains (Table 4). The main findings of this analysis are the presence of 8 LR genes in strains 1F8CT and the occurrence of self-immunity elements in four strains which confer resistance to their own bacteriocin (Table 4). On the contrary,

435 MTH17CL396 is the only strain without genes directly correlated to lantibiotics. Moreover, genes
436 encoding putative pore-forming proteins were found in five out of eight strains (Table 4). These
437 evidences support the idea that the identified genes may be involved in different functions that could
438 also be possibly related to quorum-sensing mechanisms. It is known that bacteriocin synthesis is
439 strictly linked to quorum-sensing regulation and their role as pheromone was previously demonstrated
440 (Renye and Somkuti, 2013). Alternatively, it could also be possible that the antibacterial substances
441 produced by the identified genes could be active against microorganisms different from those tested
442 in this study (Rossi et al., 2013).

443 **4. Conclusion**

444 In the present study a comparative analysis of 17 *S. thermophilus* genomes was performed and the
445 influence of geographical origin on genetic variability was assessed. Results indicated that strains
446 isolated in the same continent infrequently cluster together. A notable finding is the high divergence
447 detected for M17PTZA496 strain, which has an enhanced genome size due to several recombination
448 events. Several genetic traits related to technological phenotypes were found in all strains, with a
449 considerable degree of genome variability determined by the presence of several SNPs. These
450 findings reflected important differences among phenotypes that were detected among strains. Strains-
451 specific functions of *Streptococcus thermophilus* were investigated and valuable information
452 regarding characters of technological relevance were obtained.

453

454 **Acknowledgements**

455

456 The research was funded in part by MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca)
457 ex-60% funds and Fondazione Cassa di Risparmio di Padova e Rovigo for funded the PhD scholarship
458 of V.V. Authors are thankful to Antonio Cattelan and Christian Andrighetto for their technical support
459 and to Cristina Sartori and Paolo Gottardo for support on statistical analyses.

460

461 **Figure legends**

462

463 **Fig 1.** Phylogenetic trees of *S.thermophilus* strains. a) Analysis performed considering a panel of 400
464 conserved proteins. Strain IDs are numbered according to their geographic origin: Europe (1), Asia
465 (2), Australia (3) and North America (4). *S. salivarius*, *S. pneumoniae* and *S. macedonicus* were used
466 as outgroups. b) Analysis performed using total SNPs extracted by whole genome alignment.

467

468 **Fig 2.** Circular genome map of M17PTZA496 strain. Forward gene COG annotations, reverse gene
469 COG annotations, %GC plot and GC skew are reported in circles from outside inwards. Laterally
470 transferred islands are highlighted as green arrows (ISL1 and ISL2), “phages, prophages” protein rich
471 region (PH) as red arrow and duplicated region (DR) as yellow arrows. In the magnification a
472 schematic representation of the duplicated region is reported. Gene abbreviations indicate: DNA-
473 directed RNA polymerase alpha subunit (DNA poly; EC 2.7.7.6), translation initiation factor 1 (TIF),
474 adenylate kinase (AK, EC 2.7.4.3), preprotein translocase secY subunit (PT, TC 3.A.5.1.1); all the
475 other features refer to ribosomal proteins.

476

477 **Fig 3.** Hierarchical cluster of strains annotations. Hierarchical clustering was performed on genes
478 identified for each functional class of the SEED subsystem. Color scale on the top of the heatmap
479 allows the identification of the gene numbers for the strains reported. The clustering on top of the
480 figure shows similarities between strains. Italian strains are colored in dark grey, other strains in light
481 grey; strains possessing genes related to osmotic stress are indicated with an asterisk.

482

483 **Fig. 4.** Strain specific features. For each strain reported in the x axis, genes were assigned to 24
484 functional categories of the SEED subsystem. The number of genes in the y axis for each functional
485 category is proportional to its parcel height.

486

487 **Fig. 5.** Strains adhesion properties. EPS operon promoter analysis (a). Results from colorimetric assay
488 (b); dotted line indicates the detection threshold. Letters above the bars correspond to the statistically
489 identified groups. Images of cells aggregation in MTH17CL396 (c) and TH985 (d) by Scanning
490 Electronic Microscope.

491

492 **Tables**

493

494 **Tab. 1.** Geographical origin and genomic information of the *S. thermophilus* strains present in
 495 GenBank database (August 2015) used in the present work.

496

Strain	Isolation matrix	Geographic origin	Genome size (Mbp)	No. of ORF	No. of CRISPRs	NCBI accession number	Reference
1F8CT	curd from Raw cow milk	Veneto, Italy	1.74	1864	25	AZTK000000000	(Treu, Vendramin, Bovo, Campanaro, & Corich, 2014a)
M17PTZA496	Fontina cheese	Valle d'Aosta, Italy	2.13	2221	4	AZJT000000000	(Treu, Vendramin, Bovo, Campanaro, Corich, et al., 2014b)
MTH17CL396	Fontina cheese	Valle d'Aosta, Italy	1.82	1935	25	AZJS000000000	(Treu, Vendramin, Bovo, Campanaro, Corich, et al., 2014b)
TH982	Buffalo mozzarella curd	Campania, Italy	1.79	1924	24	AZTL000000000	(Treu, Vendramin, Bovo, Campanaro, & Corich, 2014a)
TH985	Buffalo mozzarella whey	Campania, Italy	1.83	1952	26	AZTM000000000	(Treu, Vendramin, Bovo, Campanaro, & Corich, 2014a)
TH1435	Raw goat milk	Friuli Venezia Giulia, Italy	1.79	1925	23	AYSG000000000	(Treu, Vendramin, Bovo, Campanaro, Corich, et al., 2014c)
TH1436	Raw goat milk	Friuli Venezia Giulia, Italy	1.79	1899	24	AYTT000000000	(Treu, Vendramin, Bovo, Campanaro, Corich, et al., 2014c)

TH1477	Raw cow milk	Veneto, Italy	1.9	1986	20	AZTJ000000000	(Treu, Vendramin, Bovo, Campanaro, & Corich, 2014a)
ASCC 1275	Dairy starter	Australia	1.85	1948	30	ASM69888v1	(Wu et al., 2014)
CNRZ1066	Yogurt	France	1.8	1918	14	ASM1184v1	(Bolotin et al., 2004)
JIM 8232	Milk	France	1.93	2076	21	ASM25339v1	(Delorme et al., 2011)
LMD-9	Industrial starter	USA	1.86	1930	21	ASM1448v1	(Hols et al., 2005)
LMG 18311	Yogurt	UK	1.8	1892	27	ASM1182v1	(Bolotin et al., 2004)
MN-BM-A02	Traditional dairy products	China	1.85	1953	29	ASM100801v1	(Shi et al., 2015)
MN-ZLW-002	Yogurt	China	1.85	1941	23	ASM26267v1	(Kang et al., 2012)
ND03	Traditional dairy products	China	1.83	1935	23	ASM18287v1	(Sun et al., 2011)
SMQ-301	Mozzarella whey	Canada	1.86	1952	24	ASM97166v1	(Labrie et al., 2015)

Tab. 2. Growth and fermentation parameters of *S.thermophilus* strains. Maximum acidification rate ($V_{\max}$) is expressed as ΔpH unit per 10^{-4} per minute, T_{pHx} refers to the time required to achieve pH 5.2 and 4.6 respectively, **: p-value<0.001.

Strain ID	Growth curve parameters			Acidification performance			
	λ (h)	μ (h ⁻¹)	log Nmax (cell/ml)	ΔpH	$V_{\max}$	$T_{\text{pH } 5.2}^{**}$ (h)	$T_{\text{pH } 4.6}^{**}$ (h)
1F8CT	2.93	0.53	8.5	-	-	-	-
M17PTZA496	1.12	0.42	10.3	0.99	29	10	-
MTH17CL396	1.16	0.36	9.1	0.98	61	16	-
TH982	2.05	0.41	9.5	1.74	33	7	13
TH985	1.46	0.35	9.9	1.26	24	12	23
TH1435	1.46	0.35	9.6	1.50	57	4	6
TH1436	1.53	0.45	9.0	1.62	52	4	7
TH1477	2.21	0.26	9.2	1.41	43	6	14

Tab. 3. Extracellular polysaccharides genes in *S. thermophilus* strains genome. GTF: glycosyltransferase family group 1 or 2, EpsA: exopolysaccharide biosynthesis transcriptional activator, EpsB: manganese-dependent protein-tyrosine phosphatase, EpsC: tyrosine-protein kinase transmembrane modulator, EpsD: tyrosine-protein kinase, CpsF: polysaccharide biosynthesis protein, CpsG: glycosyl transferase, CpsM: polysaccharide biosynthesis protein. If not differently specified, the percentage of identity for gene clustering by amino acidic sequences similarity was 100%.

Strain ID	GTF1	GTF2	Operon "core"					Second operon		
			EpsA	EpsB	EpsC	EpsD	EpsE	CpsF	CpsG	CpsM
1F8CT		1	1	1	1	1	1	1	1	1
M17PTZA496		1	1	1	1	1	1;1(80%)	1	1	
MTH17CL396	1		1	1	1	1	1			
TH982	1	1	1	1;1(80%)	1;1(90%)	1;1(90%);1(80%)	1			1
TH985			1	1;1(80%)	1(90%);1(80%)	1(90%);1(80%)	1			
TH1435			1	1	1	1	1			
TH1436			1	1	1	1	1			
TH1477	1		1	1	1	1	1			

Tab. 4. Genes encoding for bacteriocins in *S. thermophilus* strains. Total numbers of genes belonging to lantibiotics, ABC transporter and non-lantibiotic compounds are reported.

Strain ID	Lantibiotic related genes	ABC transporter	Lactococcin LcnD-like	Pore-forming peptide	Bacteriocin self- immunity protein
1F8CT	8	3	1	0	0
M17PTZA496	1	2	1	0	1
MTH17CL396	0	5	1	1	0
TH982	4	2	0	0	0
TH985	2	1	1	2	0
TH1435	2	4	1	2	1
TH1436	3	2	1	1	1
TH1477	1	2	1	1	1

517 Bibliography

- 518 Andrichetto, C., Borney, F., Barmaz, A., Stefanon, B., Lombardi, A. (2002). Genetic diversity of
519 *Streptococcus thermophilus* strains isolated from Italian traditional cheeses. International Dairy
520 Journal, 12, 141–144.
- 521 Awad, S., Hassan, A. N., & Muthukumarappan, K. (2005). Application of exopolysaccharide-
522 producing cultures in reduced-fat Cheddar cheese: texture and melting properties. Journal of
523 Dairy Science, 88(12), 4204–4213. doi:10.3168/jds.S0022-0302(05)73106-4
- 524 Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., ... Zagnitko, O. (2008).
525 The RAST server : rapid annotations using subsystems technology. BMC Genomics, 15, 1–15.
526 doi:10.1186/1471-2164-9-75
- 527 Baranyi, J., & Roberts, T. a. (1994). A dynamic approach to predicting bacterial growth in food.
528 International Journal of Food Microbiology, 23(3-4), 277–294. doi:10.1016/0168-
529 1605(94)90157-0
- 530 Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S. D., Kulakauskas, S., ... Hols, P. (2004).
531 Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus*
532 *thermophilus*. Nature Biotechnology, 22(12), 1554–1558. doi:10.1038/nbt1034
- 533 Caggianiello, G., Kleerebezem, M., Spano, G. (2016). Exopolysaccharides produced by lactic acid
534 bacteria: from health-promoting benefits to stress tolerance mechanisms. Applied Microbiology
535 and Biotechnology, 100(9), 3877–3886. doi:10.1007/s00253-016-7471-2
- 536 Campanaro, S., Treu, L., Vendramin, V., Bovo, B., Giacomini, A., Corich, V. (2014). Metagenomic
537 analysis of the microbial community in fermented grape marc reveals that *Lactobacillus*
538 *fabifermentans* is one of the dominant species: Insights into its genome structure. Applied
539 Microbiology and Biotechnology, 98(13), 6015–6037. doi:10.1007/s00253-014-5795-3
- 540 Chang, O. K., Perrin, C., Galia, W., Saulnier, F., Miclo, L., Roux, E., ... Dary, A. (2012). Release of
541 the cell-envelope protease PrtS in the growth medium of *Streptococcus thermophilus* 4F44.
542 International Dairy Journal, 23(2), 91–98. doi:10.1016/j.idairyj.2011.10.014
- 543 Couvigny, B., Thériat, C., Gautier, C., Renault, P., Briandet, R., & Guédon, E. (2015). *Streptococcus*
544 *thermophilus* Biofilm Formation: A Remnant Trait of Ancestral Commensal Life?. PloS one,
545 10(6), e0128099.
- 546 Dandoy, D., Fremaux, C., De Frahan, M. H., Horvath, P., Boyaval, P., Hols, P., Fontaine, L. (2011).
547 The fast milk acidifying phenotype of *Streptococcus thermophilus* can be acquired by natural
548 transformation of the genomic island encoding the cell-envelope proteinase PrtS. In Microbial
549 Cell Factories, 10, S21. doi:10.1186/1475-2859-10-S1-S21

550 Darling, A. C. E., Mau, B., Blattner, F. R., Perna, N. T. (2004). Mauve: multiple alignment of
 551 conserved genomic sequence with rearrangements. *Genome Research*, (14), 1394–1403.
 552 doi:10.1101/gr.2289704.tion

553 Delorme, C., Bartholini, C., Luraschi, M., Pons, N., Loux, V., Almeida, M., ... Renault, P. (2011).
 554 Complete genome sequence of the pigmented *Streptococcus thermophilus* strain JIM8232.
 555 *Journal of Bacteriology*, 193(19), 5581–5582. doi:10.1128/JB.05404-11

556 Divya, J. B., Nampoothiri, K. M. (2015). Folate fortification of skim milk by a probiotic *Lactococcus*
 557 *lactis* CM28 and evaluation of its stability in fermented milk on cold storage. *Journal of Food*
 558 *Science and Technology*, 52(6), 3513–3519. doi:10.1007/s13197-014-1406-7

559 Egan, K., Field, D., Rea, M. C., Ross, R. P., Hill, C., Cotter, P. D. (2016). Bacteriocins: novel
 560 solutions to age old spore-related problems? *Frontiers in Microbiology*, 7, 461.
 561 doi:10.3389/fmicb.2016.00461

562 Erkus, O., Okuklu, B., Yenidunya, A. F., Harsa, S. (2014). High genetic and phenotypic variability
 563 of *Streptococcus thermophilus* strains isolated from artisanal Yuruk yoghurts. *LWT - Food*
 564 *Science and Technology*, 58(2), 348–354. doi:10.1016/j.lwt.2013.03.007

565 Flemming, H., Wingender, J. (2010). The biofilm matrix. *Nature Reviews Microbiology*, 8(9), 623–
 566 633. doi:10.1038/nrmicro2415

567 Gaden, E. J., Bokanga, M., Harlander, S., Hesseltine, C., Steinkraus, K. (1992). Applications of
 568 biotechnology in traditional fermented foods. National Academies Press, Washington.

569 Ge, R., Sun, X., He, Q. (2009). Iron acquisition by *Streptococcus* species: an updated review.
 570 *Frontiers of Biology in China*, 4(4), 392–401. doi:10.1007/s11515-009-0035-4

571 Giraffa, G., Paris, A., Valcavi, L., Gatti, M., Neviani, E. (2001). Genotypic and phenotypic
 572 heterogeneity of *Streptococcus thermophilus* strains isolated from dairy products. *Journal of*
 573 *Applied Microbiology*, 91(5), 937–943. doi:10.1046/j.1365-2672.2001.01464.x

574 Goh, Y. J., Goin, C., O’Flaherty, S., Altermann, E., Hutkins, R. (2011). Specialized adaptation of a
 575 lactic acid bacterium to the milk environment: the comparative genomics of *Streptococcus*
 576 *thermophilus* LMD-9. *Microbial Cell Factories*, 10, S22. doi:10.1186/1475-2859-10-S1-S22

577 Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., ... Kleerebezem,
 578 M. (2005). New insights in the molecular biology and physiology of *Streptococcus thermophilus*
 579 revealed by comparative genomics. *FEMS Microbiology Reviews*, 29 (3), 435–463.
 580 doi:10.1016/j.femsre.2005.04.008

581 Huang, L. (2008). Growth kinetics of *Listeria monocytogenes* in broth and beef frankfurters--
 582 determination of lag phase duration and exponential growth rate under isothermal conditions.
 583 *Journal of Food Science*, 73(5), E235–E242. doi:10.1111/j.1750-3841.2008.00785.x

584 Huang, L. (2010). Growth kinetics of *Escherichia coli* O157:H7 in mechanically-tenderized beef.
585 International Journal of Food Microbiology, 140(1), 40–48.
586 doi:10.1016/j.ijfoodmicro.2010.02.013

587 ISTAT (2014). Annuario statistico italiano. doi:10.1017/CBO9781107415324.004

588 Iyer, R., Tomar, S. K., Maheswari, T. U., Singh, R. (2010). *Streptococcus thermophilus* strains:
589 Multifunctional lactic acid bacteria. International Dairy Journal, 20(3), 133–141.
590 doi:10.1016/j.idairyj.2009.10.005

591 Kang, X., Ling, N., Sun, G., Zhou, Q., Zhang, L., Sheng, Q. (2012). Complete genome sequence of
592 *Streptococcus thermophilus* strain MN-ZLW-002. Journal of Bacteriology, 194(16), 4428–9.
593 doi:10.1128/JB.00740-12

594 Kongo, J. M. (2013). Lactic acid bacteria as starter-cultures for cheese processing: past , present and
595 future developments., in: Marcelino Kongo (Eds.) In Lactic acid bacteria– research &
596 development for food, health and livestock purposes animal. InTech Chapters published, pp. 3–
597 22.

598 Krastel, K., Senadheera, D. B., Mair, R., Downey, J. S., Goodman, S. D., Cvitkovitch, D. G. (2010).
599 Characterization of a glutamate transporter operon, glnQHMP, in *Streptococcus mutans* and its
600 role in acid tolerance. Journal of Bacteriology, 192(4), 984–993. doi:10.1128/JB.01169-09

601 Labrie, S. J., Tremblay, D. M., Plante, P., Wasserscheid, J., Dewar, K., Corbeil, J., Moineau, S.
602 (2015). Complete genome sequence of *Streptococcus thermophilus* SMQ-301, a model strain
603 for phage-host interactions. Genome Announcements, 3(3), 12–13.
604 doi:10.1128/genomeA.00480-15.

605 Li, W., Bian, X., Evivie, S. E., & Huo, G. C. (2016). Comparative Analysis of Clustered Regularly
606 Interspaced Short Palindromic Repeats (CRISPR) of *Streptococcus thermophilus* St-I and its
607 Bacteriophage-Insensitive Mutants (BIM) Derivatives. Current microbiology, 73(3), 393-400.

608 Li, W., Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein
609 or nucleotide sequences. Bioinformatics, 22(13), 1658–1659. doi:10.1093/bioinformatics/btl158

610 Maragkoudakis, P. A., Nardi, T., Bovo, B., D’Andrea, M., Howell, K. S., Giacomini, A., Corich, V.
611 (2013). Biodiversity, dynamics and ecology of bacterial community during grape marc storage
612 for the production of grappa. International Journal of Food Microbiology, 162(2), 143–151.
613 doi:10.1016/j.ijfoodmicro.2013.01.005

614 Mora, D., Fortina, M. G., Parini, C., Ricci, G., Gatti, M., Giraffa, G., Manachini, P. L. (2002). Genetic
615 diversity and technological properties of *Streptococcus thermophilus* strains isolated from dairy
616 products. Journal of Applied Microbiology, 93, 278–287.

617 Morandi, S., Brasca, M. (2012). Safety aspects, genetic diversity and technological characterisation

618 of wild-type *Streptococcus thermophilus* strains isolated from north Italian traditional cheeses.
619 Food Control, 23(1), 203–209. doi:10.1016/j.foodcont.2011.07.011

620 Morris, L. S., Evans, J., Marchesi, J. R. (2012). A robust plate assay for detection of extracellular
621 microbial protease activity in metagenomic screens and pure cultures. Journal of
622 Microbiological Methods, 91(1), 144–146. doi:10.1016/j.mimet.2012.08.006

623 Pandey, A., Bringel, F., Meyer, J. M. (1994). Iron requirement and search for siderophores in lactic
624 acid bacteria. Applied Microbiology and Biotechnology, 40(5), 735–739.
625 doi:10.1007/BF00173337

626 R Development Core Team, R. (2008). R: a language and environment for statistical computing. in:
627 ed R. D. C. Team. Vienna. doi:10.1007/978-3-540-74686-7

628 Rasmussen, T. B., Danielsen, M., Ondrej, V., Garrigues, C., Johansen, E., & Pedersen, M. B. (2008).
629 *Streptococcus thermophilus* core genome: comparative genome hybridization study of 47
630 strains. Applied and Environmental Microbiology, 74(15), 4703. doi:10.1128/AEM.00132-08

631 Renye, J. A., Somkuti, G. A. (2013). BlpC-regulated bacteriocin production in *Streptococcus*
632 *thermophilus*. Biotechnology Letters, 35(3), 407–412. doi:10.1007/s10529-012-1095-0

633 Rossi, F., Marzotto, M., Cremonese, S., Rizzotti, L., Torriani, S. (2013). Diversity of *Streptococcus*
634 *thermophilus* in bacteriocin production: inhibitory spectrum and occurrence of thermophilin
635 genes. Food Microbiology, 35(1), 27–33. doi:10.1016/j.fm.2013.02.006

636 Russo, P., Capozzi, V., Arena, M. P., Spadaccino, G., Dueñas, M. T., López, P., ... Spano, G. (2014).
637 Riboflavin-overproducing strains of *Lactobacillus fermentum* for riboflavin-enriched bread.
638 Applied Microbiology and Biotechnology, 98, 3691–3700. doi:10.1007/s00253-013-5484-7

639 Saeed, A., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., ... Quackenbush, J. (2003). TM4: a
640 free, open-source system for microarray data management and analysis. Biotechniques, 34(2),
641 374–348.

642 Segata, N., Bornigen, D., Xochitl, M., Huttenhower, C. (2013). PhyloPhlAn is a new method for
643 improved phylogenetic and taxonomic placement of microbes. Nature Communication, (4),
644 2304. doi:10.1016/j.micinf.2011.07.011.

645 Shi, Y., Chen, Y., Li, Z., Yang, L., Chen, W., Mu, Z. (2015). Complete Genome Sequence of
646 *Streptococcus thermophilus* MN-BM- A02 , a Rare Strain with a High Acid-Producing Rate and
647 Low Post- Acidification Ability. Genome Announcements, 3(5), 2–3.
648 doi:10.1128/genomeA.00979-15.

649 Sleator, R. D., Hill, C. (2001). Bacterial osmoadaptation: the role of osmolytes in bacterial stress and
650 virulence. FEMS Microbiology Reviews, 26(1), 49–71. doi:10.1016/S0168-6445(01)00071-7

651 Snel, B., Bork, P., Huynen, M. A. (2002). Genomes in flux: the evolution of archaeal and

652 proteobacterial gene content. *Genome Research*, 12(1), 17–25. doi:10.1101/gr.176501

653 Stingle, F., Neeser, J., Mollet, B. (1996). Identification and Characterization of the eps (

654 Exopolysaccharide) Gene cluster from *Streptococcus thermophilus* Sfi6. *Journal of*

655 *Bacteriology*, 178(6), 1680–1690.

656 Sun, Z., Chen, X., Wang, J., Zhao, W., Shao, Y., Wu, L., ... Chen, W. (2011). Complete genome

657 sequence of *Streptococcus thermophilus* strain ND03. *Journal of Bacteriology*, 193(3), 793–794.

658 doi:10.1128/JB.01374-10

659 Treu, L., Toniolo, C., Nadai, C., Sardu, A., Giacomini, A., Corich, V., Campanaro, S. (2014). The

660 impact of genomic variability on gene expression in environmental *Saccharomyces cerevisiae*

661 strains. *Environmental Microbiology*, 16(5), 1378–1397. doi:10.1111/1462-2920.12327

662 Treu, L., Vendramin, V., Bovo, B., Campanaro, S., Corich, V. (2014a). Genome Sequences of Four

663 Italian *Streptococcus thermophilus* Strains. *Genome Announcements*, 2(2), 2008–2009.

664 doi:10.1128/genomeA.00126-14.

665 Treu, L., Vendramin, V., Bovo, B., Campanaro, S., Corich, V., Giacomini, A. (2014b). Genome

666 Sequences of *Streptococcus thermophilus* Strains MTH17CL396 and M17PTZA496 from

667 Fontina, an Italian PDO Cheese. *Genome Announcements*, 2(1), 2008–2009.

668 doi:10.1128/genomeA.00067-14

669 Treu, L., Vendramin, V., Bovo, B., Campanaro, S., Corich, V., Giacomini, A. (2014c). Whole-

670 Genome Sequences of *Streptococcus thermophilus* Strains TH1435 and TH1436, Isolated from

671 Raw Goat Milk. *Genome Announcements*, 2(1), e01129–13. doi:10.1128/genomeA.01129-13

672 Treu, L., Vendramin, V., Bovo, B., Giacomini, A., Corich, V., & Campanaro, S. (2014d). Genome

673 sequence of *Lactobacillus fabifermentans* strain T30PCM01, isolated from fermenting grape

674 marc. *Genome announcements*, 2(1), e00060-14.

675 Tsou, C. C., Chiang-Ni, C., Lin, Y. S., Chuang, W. J., Lin, M. T., Liu, C. C., Wu, J. J. (2010).

676 Oxidative stress and metal ions regulate a ferritin-like gene, dpr, in *Streptococcus pyogenes*.

677 *International Journal of Medical Microbiology*, 300(4), 259–264.

678 doi:10.1016/j.ijmm.2009.09.002

679 Tuimala, J. (2005). A primer to phylogenetic analysis using the PHYLIP package, fourth (ed.) Center

680 for Scientific Computing Ltd., Espoo.

681 van Heel, A. J., de Jong, A., Montalban-Lopez, M., Kok, J., Kuipers, O. P. (2013). BAGEL3:

682 automated identification of genes encoding bacteriocins and (non-)bactericidal

683 posttranslationally modified peptides. *Nucleic Acids Research*, 41, W448–W453.

684 doi:10.1093/nar/gkt391

685 Vernikos, G., Parkhill, J. (2006). Genome analysis interpolated variable order motifs for identification

686 of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. Bioinformatics,
687 22(18), 2196–2203. doi:10.1093/bioinformatics/btl369

688 Vuyst, L. De, Weckx, S., Ravyts, F., Herman, L., Leroy, F. (2011). New insights into the
689 exopolysaccharide production of *Streptococcus thermophilus*. International Dairy Journal,
690 21(9), 586–591. doi:10.1016/j.idairyj.2011.03.016

691 Willey, J. M., van der Donk, W. A. (2007). Lantibiotics: peptides of diverse structure and function.
692 Annual Review of Microbiology, 61, 477–501. doi:10.1146/annurev.micro.61.080706.093501

693 Wu, Q., Tun, H. M., Leung, F. C., Shah, N. P. (2014). Genomic insights into high exopolysaccharide-
694 producing dairy starter bacterium *Streptococcus thermophilus* ASCC 1275. Scientific Reports,
695 4(4974), 1–8. doi:10.1038/srep04974

696 Zwietering, M. H., Jongenburger, I., Rombouts, F. M., Van 't Riet, K. (1990). Modeling of the
697 Bacterial Growth Curve Modeling of the Bacterial Growth Curve. Applied and Environmental
698 Microbiology, 56(6), 1875–1881.

699