Food Microbiology 63 (2017) 47–57

Contents lists available at ScienceDirect

Food Microbiology

journal homepage: www.elsevier.com/locate/fm

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

Veronica Vendramin [a], Laura Treu [a, b], 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

ARTICLE INFO

Article history:
Received 1 July 2016
Received in revised form
27 September 2016
Accepted 1 November 2016
Available online 3 November 2016

Keywords:
Dairy products
Streptococcus thermophilus
Bacterial biodiversity
Genome comparison

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

1. Introduction

Streptococcus thermophilus is a thermophilic Lactic Acid Bacterium (LAB) of major importance in the dairy industry. Due to its ability to rapidly ferment lactose, it is widely used as starter to obtain fermented milk products, contributing to milk acidification and organoleptic properties enrichment (Giraffa et al., 2001). The long history of safe use in food production allowed *S. thermophilus* to obtain the status of Generally Recognized as Safe (GRAS) and of Qualified Presumption of Safety (QPS). At present, it is considered the second most important species of industrial LAB after *Lactococcus lactis*. It was estimated that over 10^{21} live cells are ingested annually leading the species to achieve a market value of 40 billion US$, approximately (Iyer et al., 2010). Similarly to other dairy microbes, *S. thermophilus* natural biodiversity decreases with its overuse of industrial starters, hence isolation and characterization of new strains becomes of great importance, since it may lead to the discovery of novel and desirable characteristics, which can fulfil industrial demands (Erkus et al., 2014). Strains analyzed in the present study are used as natural starters for Protected Designation Origin (PDO) Italian cheeses, i.e. Fontina, Grana Padano and Mozzarella. These products, obtained from traditional back-slopping procedures, allow the maintenance of the microbiota present in the environment where they are produced. Considering the number of Italian cheese factories involved in dairy productions (ISTAT, 2014), the Italian dairy microbiota can be considered a potential important source of new *S. thermophilus* strains. Nonetheless, until now such biodiversity has been explored only partially both from the genetic and phenotypic point of view (Andrighetto et al., 2002; Morandi and Brasca, 2012). Thanks to the next generation sequencing technology, whole *S. thermophilus* genome

* Corresponding author.
E-mail address: viviana.corich@unipd.it (V. Corich).

http://dx.doi.org/10.1016/j.fm.2016.11.002
0740-0020/

sequences are publicly available. Such information allowed to study more in depth the genetic structure of many metabolic activities of the species, such as amino acid metabolism (Hols et al., 2005), arrangement of the proteolytic system (Hols et al., 2005), resistance to bacteriophage (Li et al., 2016), biosynthesis of folate (Iyer et al., 2010), metabolism of urea (Mora et al., 2004) and biofilm formation (Couvigny et al., 2015).

More generally, *S. thermophilus* genomes have been so far analyzed and compared with other related species, specifically with pathogenic streptococci (Hols et al., 2005). Within this framework, Rasmussen et al. (2008) used microarrays to perform a comparative genomic analysis of different *S. thermophilus* strains to demonstrate the presence of variable subsets of genes responsible for ecological and technological differences. One of the most interesting technological properties reported in comparative studies is related to the ability to synthesize extracellular polymeric substances (EPS) (Flemming and Wingender, 2010; Mora et al., 2002). Recently, the beneficial effects of EPS in fermented milk have been recognized and linked to their role as thickeners and stabilizers of the product, together with healthy effects, such as their activity on human blood pressure and gastrointestinal tract health (Caggianiello et al., 2016). To date, 28 different EPS gene clusters are known in *S. thermophilus* (Iyer et al., 2010; Vuyst et al., 2011; Wu et al., 2014).

This species has been tested as bio-preservative to control growth of pathogenic and spoilage bacteria in dairy products by production of bacteriocins (Kongo, 2013). These molecules, produced by food grade lactic acid bacteria, are classified into two classes based on their modification status. Known *Streptococcus* bacteriocins belong to class I and class IIb (Egan et al., 2016) and the identification of genes encoding for bacteriocins is a quite difficult task due to their small size and high variability in sequence composition (Willey and van der Donk, 2007).

The present study used the results obtained thanks to the most innovative Next Generation Sequencing approach (Treu et al., 2014a, 2014b, 2014c) to provide more details on the genetic organization of *S. thermophilus* at whole genome level. The overall biodiversity of Italian *S. thermophilus* was studied by comparing eight isolates, coming from PDO cheese productions. It is known that subsets of features specifically characterizing different strains are extremely important when they are forced to face environmental changes (Hols et al., 2005). The geographical effect on biodiversity was examined by comparing the genome sequences of all strains available in the literature. Furthermore, technological characters related to food production were investigated in the Italian strains and linked to genomic data.

2. Material and methods

2.1. Strain used

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

2.2. Phylogenetic and genomic analyses

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

2.3. Gene content evaluation

For each *S. thermophilus* strains, annotated gene were attributed to subsystem and a features list was created based on subsystem gene abundance. Feature lists were used to elaborate hierarchical clustering (HCL) using MeV (MultiExperiment Viewer) software (Saeed et al., 2003). Strains functional relationship was computed using the "linkage method" process for determining cluster-to-cluster distances and the "Euclidean distance" for distance calculation. A comparison on subsystem gene abundance was conducted analyzing the resulting heatmap and the most interesting subsystems were manually investigated in detail. For the specific subsystem 'Iron acquisition and metabolism', the following strains of *Streptococcus pyogenes* were used: M1GAS, MGAS10270, MGAS10394, MGAS1075, MGAS2096, MGAS315, MGAS5005, MGAS6180, MGAS8232, MGAS9429, SSI-1 and str. Manfredo. *Lactobacillus fabifermentans* T30PCM01 genome was also included in the analysis (Treu et al., 2014d). On the basis of features lists, non-redundant common and non-common strain features were identified using R software, custom script (R Development Core Team, 2008). Strain specific features were assigned to the SEED categories using RAST.

2.4. Identification of sequences related to technological properties

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

Exopolysaccharides-related genes were analyzed considering subsystems completeness and sequence similarities. Genes assigned to "EPS" and "CPS" were identified for each strain, their number of copies and the organization into operons were recorded. Sequences were clustered using CD-HIT at 50%, 80%, 90% and 100% of identity in order to understand their degree of similarity.