International Journal of Food Microbiology 242 (2017) 70–81

Contents lists available at ScienceDirect

International Journal of Food Microbiology

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

Study of *Streptococcus thermophilus* population on a world-wide and historical collection by a new MLST scheme

Christine Delorme [a,*], Nicolas Legravet [a], Emmanuel Jamet [b], Caroline Hoarau [b], Bolotin Alexandre [a], Walid M. El-Sharoud [c], Mohamed S. Darwish [c], Pierre Renault [a]

[a] Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France
[b] Actilait, 419, route des champs laitiers, F-74 800 La Roche sur Foron, France
[c] Food Safety and Microbial Physiology Laboratory, Dairy Department, Faculty of Agriculture, Mansoura University, Mansoura, Egypt

ARTICLE INFO

Article history:
Received 1 August 2016
Received in revised form 16 November 2016
Accepted 17 November 2016
Available online 18 November 2016

Keywords:
Streptococci
MLST
Genetic variability
HGT
Traditional cheese

ABSTRACT

We analyzed 178 *Streptococcus thermophilus* strains isolated from diverse products, from around the world, over a 60-year period with a new multilocus sequence typing (MLST) scheme. This collection included isolates from two traditional cheese-making sites with different starter-use practices, in sampling campaigns carried out over a three years period. The nucleotide diversity of the *S. thermophilus* population was limited, but 116 sequence types (ST) were identified. Phylogenetic analysis of the concatenated sequences of the six housekeeping genes revealed the existence of groups confirmed by eBURST analysis. Deeper analyses performed on 25 strains by CRISPR and whole-genome analysis showed that phylogenies obtained by MLST and whole-genome analysis were in agreement but differed from that inferred by CRISPR analysis. Strains isolated from traditional products could cluster in specific groups indicating their origin, but also be mixed in groups containing industrial starter strains. In the traditional cheese-making sites, we found that *S. thermophilus* persisted on dairy equipment, but that occasionally added starter strains may become dominant. It underlined the impact of starter use that may reshape *S. thermophilus* populations including in traditional products. This new MLST scheme thus provides a framework for analyses of *S. thermophilus* populations and the management of its biodiversity.

1. Introduction

Streptococcus thermophilus is a thermophilic lactic acid bacterium (LAB) originating from traditional dairy products for which a thermic step is implemented in the process (Sheehan et al., 2007). This species, generally recognized as safe for food products and of major economic importance, is systematically used with *Lactobacillus delbrueckii* subsp. *bulgaricus* in yogurt manufacture. Nowadays, it is also used in cheese production, alongside other LAB, such as *Lactococcus lactis* (Hols et al., 2005) and various lactobacilli (Giraffa et al., 2010). Isolates of these populations have been selected for a variety of interesting properties and largely industrialized in commercial starters (Hassan and Frank, 2001). Moreover, *S. thermophilus* strains are often isolated from traditional and artisanal cheese production systems around the world. *S. thermophilus* are particularly widespread in cheeses of specific territories as in France (Duthoit et al., 2003), Italy (Carraro et al., 2011; Dolci et al., 2010; Lazzi et al., 2009; Lortal et al., 2009; Randazzo et al., 2010), Spain (Alegria et al., 2009), Lebanon (Serhan et al., 2009), Turkey (fermented food) (Sengun et al., 2009), Slovakia (Bryndza cheese) (Chebenova-Turcovska et al., 2011), Greece (Graviera cheese) (Samelis et al., 2011) and Egypt and Algeria (fermented milk) (El-Baradei et al., 2008; El-Sharoud et al., 2011; Mezaini et al., 2009). Interestingly, these products are manufactured by traditional processes where the development of flora is entirely dependent on the bacteria present in the farm, and/or of procedures such as backslopping. These natural strains, also named wild, indigenous, autochthonous, local or original isolates may be specific to these products and/or geographic area by contrast to well-defined industrial starter strains largely widespread in dairy business (Montel et al., 2014). Natural strains may present original features leading to an increasing interest in the analysis of their genome to explore new technological or probiotic properties (Kang et al., 2012; Prajapati et al., 2013; Sun et al., 2011; Treu et al., 2014a, 2014b, 2014c; Wu et al., 2014).

The analysis of *S. thermophilus* genomes provided new insight into the evolution of this species (Bolotin et al., 2004; Goh et al., 2011; Hols et al., 2005). The high proportion of pseudogenes (10%) in these genomes indicates that this bacterium has evolved through a process involving a loss of functions. In addition, recombination and gene acquisition by horizontal gene transfer (HGT) have also contributed to the shaping of *S. thermophilus* genome and its plasticity (Bolotin et al., 2004; Delorme et al., 2010; Lefebure and Stanhope, 2007; Liu et al.,

* Corresponding author.
E-mail address: christine.delorme@inra.fr (C. Delorme).

http://dx.doi.org/10.1016/j.ijfoodmicro.2016.11.016
0168-1605/

2009; Marri et al., 2006). More than 20 *S. thermophilus* genome sequences are now available, for industrial starters (Bolotin et al., 2004; Makarova et al., 2006) and potential natural strains isolated from traditionally fermented milk in different countries (Kang et al., 2012; Labrie et al., 2015; Prajapati et al., 2013; Shi et al., 2015; Sun et al., 2011; Treu et al., 2014a, 2014b, 2014c; Wels et al., 2015; Wu et al., 2014). Intriguingly, comparative genomic analyses have shown the genome of *S. thermophilus* strain NDO3 isolated from naturally fermented yak milk in Qinghai (China) is very similar to those of industrial starter strains, such as CNRZ1066, LMG18311 and LMD-9 (Sun et al., 2011). Indeed, several studies support that the modification of the cheese manufacturing practices by the widespread use of industrial starters worldwide over the last few decades may have shaped LAB populations and reduce diversity in dairy products (Beuvier and Buchin, 2004; Rahman et al., 2014).

In recent years, the diversity of the lactic acid bacteria present in food ecosystems has been studied in detail by various molecular fingerprinting techniques, such as RAPD-PCR and PFGE (pulsed-field gel electrophoresis), which have provided general information about the diversity of several species, including *S. thermophilus* (Ercolini et al., 2005; Giraffa et al., 2001; Mora et al., 2002; Moschetti et al., 1998). Finally, the use of clustered regular interspaced short palindromic repeats (CRISPR) has been proposed for typing and comparative analyses of strains in *S. thermophilus* (Horvath et al., 2008). Several ways of using CRISPR have been developed and have proved useful for subtyping, for monitoring pathogen outbreaks [see for a review (Shariat et al., 2013)]. In a previous study, a MLST scheme was developed for the closely related bacterium *Streptococcus salivarius* (Delorme, 2008). This scheme was then extended to a set of *S. thermophilus* isolates, for investigation of the phylogenic relationships of *S. thermophilus* within the *salivarius* group (Delorme et al., 2010). However, the data obtained indicated that *S. thermophilus* presented a small pool of unique alleles with low sequence diversity (mean of 0.2%). Recently, a MLST scheme was developed with 10 genes, but still displaying low divergence, to analyze Chinese strains (Yu et al., 2015). Indeed, this species tends to be rather uniform genetically. The low level of divergence of alleles in *S. thermophilus* probably reflects the recent emergence of the *S. thermophilus* population, resulting in a relatively clonal structure (Delorme et al., 2010). It was possible to type strains with this MLST scheme, but the small number of variable positions considerably limited the accuracy of this typing, and, therefore, its use for the definition of population structure.

The aim of the present study is to gain a better understanding of *S. thermophilus* population diversity and the potential effect of inoculation practices on its structure. For this purpose, we developed a more discriminating MLST scheme with housekeeping genes displaying higher divergence for *S. thermophilus* allowing thus an improved assessment of the phylogenic relationships between *S. thermophilus* isolates. This scheme was applied to a large collection of strains collected from around the world, over a 60-year period. These strains originated from various products, including traditional yogurt and cheeses from France, Italy, India, Greece and Egypt. A clustering analysis based on this scheme distinguished clusters of natural strains associated with specific origins, which are potential sources of new starters with interesting technological properties. Finally, our study revealed the impact of starter use on *S. thermophilus* populations, including those in traditional products. Our study provides a tool for ecological studies of *S. thermophilus* and to manage the selection of new strains.

2. Materials and methods

2.1. Screening for S. thermophilus in traditional cheese making environments

Strains were isolated from two different French sites, which use traditional processes to produce hard and semi-hard cooked cheeses, in two sampling campaigns at one site and three at the other. In the first process (PA), raw cow's milk was collected and directly processed into cheese on the farm site, by pasteurization and curdling of the milk at 38 °C. In the second process (PB), cheese was produced from milk collected from different farms and the fermentation process involved the inoculation of the pasteurized cow's milk with a fresh whey culture obtained from the previous day's cheese-making activities, with curdling at 54 °C. For both processes, we collected 35 to 40 samples in each campaign, from the farm area, including grass, cow stool, dust and litter, udder, water and fresh raw milk samples, and from the cheese production site, including the ground, milk tank, stored milk, pipes, utensils (mold, table, vats, thermometer, wooden tools, stirring blades and axis), rennet and the cheese itself at 1 h, and 20 h after folding and after three months of ripening. The sampling campaigns were carried out in January and July, corresponding to periods during which the cows were housed indoors (January) or were outdoors in mountain pastures, at elevations of 1500–1800 m (July). Strains were isolated by plating diluted samples from “grass to cheese”, on M17lac medium supplemented with thallium acetate. Clones picked from these plates were then tested on KF Streptococcus Broth (isolation of *enterococci*; Difco Laboratories, Detroit, MI, USA), MRS (isolation of *lactobacilli*; Merck, Darmstadt, Germany), and Chapman (isolation of *staphylococci*; Difco Laboratories, Detroit, MI, USA) medium plates, to eliminate those corresponding to *enterococci*, *lactobacilli*, *Leuconostoc*, *pediococci*, *corynebacteria*, *staphylococci* and *micrococci*. Clones corresponding to *S. thermophilus* were checked by PCR, with specific primers targeting *serB* (Table S1; El-Sharoud et al., 2011). In PA, we screened 5000 potential clones from 80 samples, to obtain 252 streptococcal isolates, and, ultimately, 115 *S. thermophilus* isolates from the two sampling campaigns. In PB, we screened 6870 clones from the 84 samples obtained in the three sampling campaigns. We obtained 356 *S. thermophilus* strains, 186 randomly selected of which were analyzed (Table S2). Finally, redundant clones, corresponding to the same strain recovered from the same sample, were eliminated by sequencing the highly variable *eps* gene promoter sequence (P*eps* region) and/or two MLST loci (*proA* and *ptsI*) (Table S1). In all cases tested in this study, *S. thermophilus* strains presenting the same *proA-ptsI-*P*eps* or *proA-ptsI* sequences were of the same sequence type (ST; data not shown). The strains were then typed by MLST.

2.2. S. thermophilus strain collection

The 178 *Streptococcus thermophilus* strains used in this MLST analysis are listed in Table 1. Isolates were obtained from various sources [starter (industrial), natural starter, yogurt, milk, cheese, whey and dairy environment] from 17 countries over a 60-year period (1950 to 2012).

The collection consisted of 12 strains representative of the ST of the isolates (PA and PB sources) recovered during this study, 49 strains from CIRM-Food bacteria (CNRZ), Rennes, France, 15 from our collection (JIM) and one from Cochin Hospital (CCH), France. We also included 17 strains recently isolated from traditional Egyptian cheeses (El-Sharoud et al., 2011). Natural strains (ITFF/ITG(ST)/EMFR/EMLN/IOFF/IOLN/IOLC) were provided by Actilait (France). Following the removal of duplicate strains on the basis of the P*eps* sequence (Table S1), a final set of 48 different natural strains was studied. Strains were grown overnight in M17 broth containing 1% lactose, at 42 °C, under an anaerobic atmosphere. The collection also contained 14 genomic DNA samples corresponding to strains from the LB Bulgaricum collection (LBB), Sofia, Bulgaria and (iii) 22 publicly available genome sequences described below.

2.3. Multilocus sequence typing

Three loci—*ddlA*, *glcK* and *tkt* (encoding D-alanine-D-alanine ligase, glucose kinase and transketolase, respectively)—used for the previous MLST analysis (Delorme et al., 2010) were retained for this study