International Journal of Food Microbiology 236 (2016) 138–147

Contents lists available at ScienceDirect

International Journal of Food Microbiology

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

Polymorphism of the phosphoserine phosphatase gene in *Streptococcus thermophilus* and its potential use for typing and monitoring of population diversity

Annamaria Ricciardi [a], Francesca De Filippis [b], Teresa Zotta [c], Angelo Facchiano [c], Danilo Ercolini [b], Eugenio Parente [d,*]

[a] Scuola di Scienze Agrarie, Forestali, Alimentari ed Ambientali, Università degli Studi della Basilicata, Potenza, Italy
[b] Department of Agricultural Sciences, Division of Microbiology, University of Naples Federico II, Portici, Italy
[c] Istituto di Scienze dell'Alimentazione, CNR, Avellino, Italy
[d] Dipartimento di Scienze, Università degli Studi della Basilicata, Potenza, Italy

ARTICLE INFO

Article history:
Received 7 April 2016
Received in revised form 7 July 2016
Accepted 26 July 2016
Available online 27 July 2016

Keywords:
Streptococcus thermophilus
Microdiversity
Phosphoserine phosphatase
High-throughput sequencing

ABSTRACT

The phosphoserine phosphatase gene (*serB*) of *Streptococcus thermophilus* is the most polymorphic gene among those used in Multilocus Sequence Typing schemes for this species and has been used for both genotyping of isolates and for evaluation of population diversity. However, the information on the potential of this gene as a marker for diversity in *S. thermophilus* species is still fragmentary. In this study, we evaluated *serB* nucleotide polymorphism and its potential impact on protein structure using data from traditional sequencing. In addition we evaluated the ability of *serB* targeted high-throughput sequencing in studying the diversity of *S. thermophilus* populations in cheese and starter cultures. Data based on traditional cultivation based techniques and sequencing provided evidence that the distribution of *serB* alleles varies significantly in some environments (commercial starter cultures, traditional starter cultures, cheese). Mutations had relatively little impact on predicted protein structure and were not found in domains that are predicted to be important for its functionality. Cultivation independent, *serB* targeted high-throughput sequencing provided evidence for significantly different alleles distribution in different cheese types and detected fluctuations in alleles abundance in a mixed strain starter reproduced by backslopping. Notwithstanding some shortcomings of this method that are discussed here, the cultivation independent approach appears to be more sensitive than cultivation based approaches based on isolation and traditional sequencing.

1. Introduction

Streptococcus thermophilus is one of the most important dairy starter species, and it is used in defined starter cultures for the production of a variety of cheeses and fermented milks, including yoghurt (Iyer et al., 2010; Parente and Cogan, 2004). It also plays a fundamental role in the production of many traditional cheeses, and its sources may be the raw milk, cheese-making equipment, and artisanal starter cultures (De Filippis et al., 2014; Montel et al., 2014; Parente et al., 2016; Quigley et al., 2013; Settanni et al., 2012). Its adaptation to the dairy environments has been reflected by gene loss and acquisition (Delorme et al., 2010; Rasmussen et al., 2008). Recently, *S. thermophilus* strains of plant origin have been described (Umamaheswari et al., 2013). Although the genome of *S. thermophilus* is relatively small (1.75–2.06 Mbp), a significant diversity in technologically relevant features has been described (Delorme et al., 2010; Iyer et al., 2010; Mora et al., 2002; Rasmussen et al., 2008; Rossi et al., 2013; Vaningelgem et al., 2004b; Zotta et al., 2008).

In traditional products obtained with artisanal starters, the diversity and dynamics of *S. thermophilus* populations are relevant for product quality and identity. In fact, the use of defined starters with limited strain diversity, for the manufacture of several PDO cheeses whose standard of identity requires the use of undefined starters, is a fraud and functional and compositional diversity of communities and populations in traditional cheeses is known to affect their sensory properties (Montel et al., 2014).

Cultivation-based approaches for the evaluation of microbial diversity in *S. thermophilus* populations have been usually based on the use of RAPD-PCR (Andrighetto et al., 2002; Mora et al., 2002), although AFLP (Lazzi et al., 2009) or PFGE (Erkuş et al., 2014; Jenkins et al., 2002; O'Sullivan and Fitzgerald, 1998) offer a better resolution. More recently, Multilocus Sequence Typing has emerged as a tool for the study of evolutionary relationships and the population structure in this species (Delorme et al., 2010; Yu et al., 2015). The relatively high polymorphism in the phosphoserine phosphatase gene (*serB*) has been exploited for genotyping of *S. thermophilus* strains isolated from Egyptian dairy

* Coresponding author at: Dipartimento di Scienze, Università degli Studi della Basilicata, Viale dell'Ateneo Lucano 10, 85100 Potenza, Italy.
E-mail address: eugenio.parente@unibas.it (E. Parente).

http://dx.doi.org/10.1016/j.ijfoodmicro.2016.07.031
0168-1605/

products (El-Sharoud et al., 2012) and from natural milk cultures (Guidone et al., 2016; Parente et al., 2016). Usually a variety of *S. thermophiles* strains can be found in traditional products, although only a few dominate in each sample (Guidone et al., 2016; Rossetti et al., 2008; Settanni et al., 2012). Even if methods based on MLST or PFGE are powerful tools in strain discrimination, culture-dependent approaches may significantly underestimate the microdiversity (i.e. the diversity below the species and subspecies level, Schloter et al., 2000) of *S. thermophilus* in dairy products. In fact, the number of strains isolated and typed for each sample is usually low (≤20) and sub-dominant strains may escape detection. In addition, a significant proportion of cells might be non-cultivable (Bottari et al., 2010).

Biotype-level detection of *S. thermophilus* with cultivation independent approaches has been achieved in some cases by PCR-DGGE (Ercolini et al., 2005; Guidone et al., 2016). However, amplicon-based high-throughput sequencing of species-specific targets offers valuable opportunities for a quantitative biotype monitoring (Ercolini, 2013) and has been exploited to study microdiversity of *S. thermophilus* population in natural starters or cheese curds (De Filippis et al., 2014; Parente et al., 2016). *serB* targeted high-throughput sequencing was previously used to evaluate the changes in the abundance of *S. thermophilus* sequence types in model natural milk cultures reproduced by back-slopping over 7–13 cycles (Parente et al., 2016). A relatively low microdiversity was found, with only 1–3 sequence types (ST) in each culture, and the dominance of few of them in all the starters. In addition, the pattern of alleles detected by cultivation-dependent and cultivation-independent methods differed. This diversity was significantly lower than that found by De Filippis et al. (2014) who used a region upstream the *lacSZ* operon as a target.

The objectives of this study were to deepen the knowledge on the polymorphism of the *serB* gene and of its gene product, phosphoserine phosphatase, in *S. thermophilus,* in order to evaluate its use for the evaluation of microdiversity of *S. thermophilus* in dairy environments through amplicon-targeted high-throughput sequencing. For this purpose we used our data and those collected from public databases to evaluate the potential occurrence of product- or geographic-specific patterns of *serB* alleles estimated by cultivation and traditional sequencing techniques. This approach was complemented by the evaluation of the distribution of *serB* sequence types in a defined strain starter reproduced by back-slopping and in a variety of cheeses by cultivation independent methods.

2. Materials and methods

2.1. Microbial strains

The microbial strains used in this study are listed in Table 1. All strains were maintained as freeze-dried stocks (in 10% reconstituted skim milk with 0.1% sodium ascorbate) in the culture collection of the Laboratory of Industrial Microbiology, Università degli Studi della Basilicata. Working cultures were routinely cultured in LM17 broth (M17 broth with 1% w/v lactose monohydrate) for 6 h at 42 °C and stored refrigerated for <1 week.

2.2. Phylogenetic analysis of phosphoserine phosphatase (serB) gene

The sequences of the complete *serB* gene of *S. thermophilus* and a number of closely related species of the genus *Streptococcus* were retrieved from complete or draft genomes available in public databases (GenBank, Integrated Microbial Genomes portal). Partial sequences of the *serB* gene were obtained in our laboratory (Table 1) as described before (Parente et al., 2016) or retrieved from the Institut Pasteur MLST and whole genome MLST database (http://bigsdb.web.pasteur.fr/streptococcus/streptococcus.html). Further sequences were retrieved using Mega BLAST searches from the NCBI non-redundant nucleotide sequence database. Sequences for other genes used in MLST for *S. thermophilus* were retrieved from either the pubmlst database (http://pubmlst.org/sthermophilus/) or from the Institut Pasteur MLST database (Supplementary Table 1). MEGA v.6 software (Tamura et al., 2013) was used to perform multiple sequence alignments (CLUSTAL algorithm) and phylogenetic analysis. The pairwise deletion for missing data/gaps was used for the calculation of pairwise distances and identification of *serB* alleles (Table 1) while the complete deletion method was used in dendrograms. DnaSP 5.10 (Librado and Rozas, 2009) was used for polymorphism analysis and Tajima neutrality test (Tajima, 1989).

Amino acid sequences were aligned using ClustalOmega (Sievers et al., 2011) at the EBI web server (http://www.ebi.ac.uk/Tools/msa/clustalo/). Results were analysed using the FASMA tool (Costantini et al., 2007) to format the alignment and evaluate percentage of identity.

2.3. Phosphoserine phosphatase protein modelling

Protein modelling was performed by applying the comparative modelling technique. Template structure was selected by means of BLAST (http://blast.ncbi.lnm.nih.gov), looking within the Protein Data Bank sequences and using the sequence of *S. thermophilus* CNRZ1066 phosphoserine phosphatase coding DNA sequence (CDS) as a query. The results of the BLAST search suggested that the structure of the phosphoserine phosphatase from *Methanococcus jannaschii* (PDB code: 1F5S) could be used for the comparative modelling (sequence identity 45%, query coverage is 97%). Protein modelling was performed by using Modeller 9.12 (Šali and Blundell, 1993). The modelling procedure used involved the generation of 10 models for each amino acid sequence. The best model for each protein was selected on the basis of energetic and structural quality evaluations, obtained by Vadar (Willard et al., 2003) and ProsaWeb (Wiederstein and Sippl, 2007). Further structural analyses and visualization were performed in Discovery Studio (DassaultSystèmes BIOVIA, Discovery Studio Modeling Environment, Release 4.5, San Diego: DassaultSystèmes, 2015).

2.4. Growth of pure and mixed cultures in milk

S. thermophilus strains with different *serB* alleles (Table 1) were reproduced with two transfers (6 h, 42 °C, followed by storage at 4 °C for ≤3 d) in reconstituted (11% w/v) skim milk (RSM, sterilized at 121 °C for 5 min) and stored at 4 °C for <16 h. The milk cultures were used to inoculate (1% v/v) pre-warmed RSM (2 × 30 mL) and incubated at 42 °C in a water bath. Samples (1 mL) were aseptically obtained at 20–30 min intervals and decimal dilutions were carried out in sterile quarter strength Ringer's solution. Bacterial counts were carried out by spiral plating on LM17 agar (LM17 broth with 1.2% bacteriological agar) using a Whitley Automated Spiral Plater II (WASP II) (50 μL inoculum, logarithmic mode). Colonies were enumerated using a digital colony counter (EasyCount 2, bioMérieux) after incubation at 42 °C for 24 h. Modelling of the growth curves was carried out using the D-model (Baranyi and Roberts, 1994). Curve fit was performed using DMFit on line (http://browser.combase.cc/DMFit.aspx). Each growth curve experiment was carried out in duplicate.

The growth kinetics of a mixed strain starter originally including six strains carrying different *serB* alleles (LA3g, serB-3; AV7_16c7, serB-30; CL13B_7a_5, serB-23c; LA1g, serB-29; N17p, serB-7b; AV6_16c7, serB-10) was also monitored. The starter had originally been assembled using equal volumes of pure milk cultures of the six strains and reproduced by back-slopping (2% of a coagulated culture was used to inoculate the following batch, and incubated for 4 h at 42 °C). The mixed culture was used to inoculate (2% v/v) pre-warmed (42 °C) RSM in Falcon tubes. Incubation, viable cell counts and primary modelling were carried out as described above. At the end of the incubation, the culture was refrigerated and used to inoculate a fresh batch of RSM on the following day, simulating the traditional back-slopping procedure. The