J. Dairy Sci. 99:1–12
http://dx.doi.org/10.3168/jds.2016-10857

Multilocus sequence typing of *Lactobacillus casei* isolates from naturally fermented foods in China and Mongolia

Qiuhua Bao, Yuqin Song, Haiyan Xu, Jie Yu, Wenyi Zhang, Bilige Menghe, Heping Zhang, and Zhihong Sun[1]
Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education, School of Food Science and Engineering, Inner Mongolia Agricultural University, Huhhot 010018, China

ABSTRACT

Lactobacillus casei is a lactic acid bacterium used in manufacturing of many fermented food products. To investigate the genetic diversity and population biology of this food-related bacterium, 224 *Lb. casei* isolates and 5 reference isolates were examined by multilocus sequence typing (MLST). Among them, 224 *Lb. casei* isolates were isolated from homemade fermented foods, including naturally fermented dairy products, acidic gruel, and Sichuan pickles from 38 different regions in China and Mongolia. The MLST scheme was developed based on the analysis of 10 selected housekeeping genes (*carB*, *clpX*, *dnaA*, *groEL*, *murE*, *pyrG*, *pheS*, *recA*, *rpoC*, and *uvrC*). All 229 isolates could be allocated to 171 unique sequence types, including 25 clonal complexes and 71 singletons. The high index of association value (1.3524) and standardized index of association value (0.1503) indicate the formation of an underlying clonal population by all the isolates. However, split-decomposition, relative frequency of occurrence of recombination and mutation, and relative effect of recombination and mutation in the diversification values confirm that recombination may have occurred, and were more frequent than mutation during the evolution of *Lb. casei*. Results from Structure analyses (version 2.3; http://pritch.bsd.uchicago.edu/structure.html) demonstrated that there were 5 lineages in the *Lb. casei* isolates, and the overall relatedness built by minimum spanning tree showed no clear relationship between the clonal complexes with either the isolation sources or sampling locations of the isolates. Our newly developed MLST scheme of *Lb. casei* was an easy and valuable tool that, together with the construction of an MLST database, will contribute to further detailed studies on the evolution and population genetics of *Lb. casei* from various niches.

Received January 5, 2016.
Accepted March 28, 2016.
[1]Corresponding author: sunzhihong78@163.com

Key words: naturally fermented foods, *Lactobacillus casei*, multilocus sequence typing, genetic diversity

INTRODUCTION

Lactic acid bacteria (**LAB**) belong to a large family of gram-positive bacteria that can ferment glucose to lactic acid as the major metabolic end product. Lactic acid bacteria are often responsible for food preservation and flavor development, and some LAB are highly beneficial to human health. *Lactobacillus casei* is an important LAB species that is important in the production of many fermented foods. It is often present in traditional fermented food products, such as cheese, qula (Bao et al., 2012a,b), koumiss (Sun et al., 2010b), kurut (Sun et al., 2010a), pickle (Yu et al., 2012), and fermented acidic gruel (Yu et al., 2011a). Most importantly, these bacteria can enter the host's gastrointestinal tract, and thus effectively confer beneficial effects within the host gut; in particular, they can improve nutrition and aid disease prevention and therapy. Thus, some *Lb. casei* isolates are regarded as probiotics and are consumed specifically because of the potentially desirable properties they confer on the host (Zhang et al., 2014).

Accurate identification and an understanding of the genetic diversity and population biology of LAB isolates is essential both for basic research in bacteriology and for continued improvement of the food industry, particularly the dairy industry. In the past, LAB species have been classified using phenotypic identification methods, particularly the sugar fermentation test (Wang et al., 2008). However, phenotypic tests often give ambiguous results. Subsequently, many researchers have been trying to discriminate them using molecular methods, such as 16S rRNA gene sequencing (Bao et al., 2012a), species-specific PCR (Bao et al., 2012a), random fragment length polymorphism (Bao et al., 2012a), SDS-PAGE profiles (Gatti et al., 2001), random amplified polymorphic DNA-PCR, and amplified fragment length polymorphism (Di Cagno et al., 2010). Although these methods have advantages, their reliability in distinguishing between similar isolates or

species is still limited, and this hinders studies on the population biology of LAB.

With the development of gene sequencing technology and a decrease in the costs associated with sequencing, multilocus sequence typing (**MLST**) is increasingly regarded as the bacterial identification and typing method of choice. Multilocus sequence typing is principally based on the partial nucleotide sequences of multiple housekeeping genes (Tanigawa and Watanabe, 2011), is highly discriminatory, and provides unambiguous results because it is based directly on nucleotide sequences. This method was first described in 1998 (Maiden et al., 1998) and several publicly accessible MLST databases are now available, including MLST (http://www.mlst.net/) and PubMLST (http://pubmlst.org/).

The MLST method has been used in several important studies on genetic diversity and bacterial pathogen population biology (Tanabe et al., 2007; Usein et al., 2014). Several MLST schemes have also been developed for species of LAB, including *Lb. casei* (Cai et al., 2007; Diancourt et al., 2007), *Lactobacillus acidophilu*s (Ramachandran et al., 2013), *Lactobacillus delbrueckii* (Tanigawa and Watanabe, 2011), *Lactobacillus plantarum* (de Las Rivas et al., 2006; Tanganurat et al., 2009), *Lactobacillus sanfranciscensis* (Picozzi et al., 2010), *Lactobacillus helveticus* (Sun et al., 2015), *Lactobacillus sakei* (Chaillou et al., 2013), *Lactococcus lactis* (Xu et al., 2014), and *Leuconostoc lactis* (Dan et al., 2014). In these studies, strains were classified with high resolution by using MLST based on the analysis of housekeeping gene sequences, including *aroE*, *ddl*, *dnaE*, *fusA*, *ftsZ*, *glnA*, *gyrB*, *glpF*, *gltX*, *gyrB*, *gpd*, *gdh*, *hemN*, *hsp60*, *ileS*, *lepA*, *leuS*, *ldhL*, *mutS*, *mutL*, *metRS*, *nrdD*, *pepV*, *pgm*, *polA*, *recA*, *recG*, *recP*, *xpt*, *yqil*, *tkt*, and *tpi*. These published reports clearly demonstrated that MLST analysis is a powerful tool for genotyping and for population biology studies of specific bacterial species.

With respect to *Lb. casei*, this species has high genotypic diversity, numerous subspecies and its population characteristics within the natural environment are unclear (Dobson et al., 2004). For this reason, we aimed to develop more effective MLST scheme for *Lb. casei* using 10 housekeeping genes on isolates from a wide range of naturally fermented foods from China and Mongolia. Such a scheme should facilitate interesting insights into the genetic diversity, population biology, and evolutionary relationships among the isolates evaluated.

MATERIALS AND METHODS

Bacterial Isolates

A total of 224 *Lb. casei* isolates were selected from the LAB Culture Collection of the Inner Mongolia Agriculture University. They had been identified using various methods including conventional phenotypic identification, 16S rRNA sequence, and species-specific PCR (Airidengcaicike et al., 2010; Yu et al., 2011a,b, 2012; Bao et al., 2012a,b; Liu et al., 2012). These included *Lb. casei* isolates from several naturally fermented dairy foods, including yogurt (fermented cow milk), kurut (fermented yak milk), koumiss (fermented mare milk), whey, cream, qula (a kind of traditional cheese), and other naturally fermented nondairy foods, such as Sichuan pickles and acidic gruel (a popular traditional fermented food in the western regions of in Inner Mongolia in China). The products from which the isolates came originated from 38 different regions in 5 provinces (Inner Mongolia, Xinjiang, Tibet, Sichuan, and Gansu) in China, and from 2 cities (Bulgan and Zavkhan) in Mongolia between 2002 and 2009. In addition, sequences of 5 reference isolates (ATCC 334, Zhang, 12A, DSM5622, and N1115) from the NCBI genome database (http://www.ncbi.nlm.nih.gov/genome/652) were also used. The whole genome of *Lb. casei* Zhang was sequenced in our laboratory (Zhang et al., 2010). Information for all isolates is listed in Supplemental Table S1 (http://dx.doi.org/10.3168/jds.2016-10857).

These *Lb. casei* isolates were freeze-dried in a cryoprotectant (10% skim milk broth containing 0.1% sodium glutamate) before being stored at −80°C before use. As a large proportion of the studied isolates (133) were from Inner Mongolia, which is a vast territory covering 700,000 km^2 in northern China, the region was subdivided into 3 zones. The first zone was Inner Mongolia A, which covered the west of the region including the cities of Bayinnaoer, Ordos, and Baotou. The second zone was Inner Mongolia B, which covered the central part of the region including the cities of Hohhot and Xilingol. The final zone was Inner Mongolia C, which covered the east of the region including the cities of Hulunbeier and Chifeng.

DNA Extraction

All *Lb. casei* isolates were retrieved from storage and each was grown in 5 mL of de Man, Rogosa, Sharpe broth (Difco, Becton Dickinson and Company, Franklin Lakes, NJ) at 37°C for 18 to 22 h without shaking. Bacteria from each isolate were harvested by centrifugation and resuspended in 500 μL of TE solution (pH 8.0, 10 m*M* Tris hydrochloride, 1 m*M* EDTA). Each suspension was put through 3 cycles of freezing and thawing in liquid nitrogen and a 65°C water bath. Genomic DNA was extracted by a modified cetyltrimethylammonium bromide (CTAB) method as described previously (Bao et al., 2012a). Purified DNA was diluted to a final concentration of 100 ng/μL for further evaluation.