J. Dairy Sci. 98:1–5
http://dx.doi.org/10.3168/jds.2014-9272
© American Dairy Science Association®, 2015.

Short communication: Single molecule, real-time sequencing technology revealed species- and strain-specific methylation patterns of 2 *Lactobacillus* strains

Wenyi Zhang, Zhihong Sun, Bilige Menghe, and Heping Zhang[1]
Key Laboratory of Dairy Biotechnology and Engineering, Education Ministry of P. R. China, Inner Mongolia Agricultural University, Hohhot, Inner Mongolia 010018, P. R. China

ABSTRACT

Pacific Biosciences' (Menlo Park, CA) single molecule, real-time sequencing technology was reported to have some advantages in generating finished genomes and characterizing the epigenome of bacteria. In the present study, this technology was used to sequence 2 *Lactobacillus* strains, *Lactobacillus casei* Zhang and *Lactobacillus plantarum* P-8. Previously, the former bacterium was sequenced by an Applied Biosystems 3730 DNA analyzer (Grand Island, NY), whereas the latter one was analyzed with Roche 454 (Indianapolis, IN) and Illumina sequencing technologies (San Diego, CA). The results showed that single molecule, real-time sequencing resulted in high-quality, finished genomes for both strains. Interestingly, epigenome analysis indicates the presence of 1 active N^6-methyladenine methyltransferase in *L. casei* Zhang, but none in *L. plantarum* P-8. Our study revealed for the first time a completely different methylation pattern in 2 *Lactobacillus* strains.
Key words: SMRT sequencing technology, epigenome, base modification

Short Communication

Lactobacillus is a group of bacteria that can ferment hexose carbohydrates to produce predominantly lactic acid. The historical discovery and exploration of these bacteria are closely associated with food fermentations, where they remain important today with industrial application value (Stiles and Holzapfel, 1997). In addition to the indigenous food-related habitats of these bacteria (Zhang et al., 2008b; Fonseca et al., 2013; Nguyen et al., 2013), they are abundant in the gastrointestinal tracts of both humans and animals (Ahrné et al., 1998; Du Toit et al., 2001). Isolates from the same host species, including *Lactobacillus plantarum* and *Lactobacillus casei*, are often detected in a broad range of ecological niches, reflecting their versatility (Kleerebezem et al., 2003; Zhang et al., 2010).

Genomics studies of *Lactobacillus* started with the publication of the *L. plantarum* WCSF1 in 2003 (Kleerebezem et al., 2003). Since then, benefitting from the quick development of DNA sequencing technology, the number of sequenced *Lactobacillus* genomes has grown exponentially. Currently, with the booming data from various genome-sequencing projects, we have unprecedented opportunities to gain insights into the genetic basis of *Lactobacillus*. Thus far, more than 182 genomes with abundant draft sequences from *Lactobacillus* are available, covering 54 different species (http://www.ncbi.nlm.nih.gov/). Although our understanding of *Lactobacillus* genetics has been significantly improved with the expanding information, their biological properties remain somewhat obscure due to the limited number of published complete genomes. Seemingly, new advances in cost-effective sequencing technologies that can help to decipher the sequenced genomes of *Lactobacillus* are still anticipated (Koren et al., 2012).

Pacific Biosciences' (Menlo Park, CA) single molecule, real-time sequencing technology (**SMRT**) was reported to have some advantages in assembling genomes and characterizing epigenome (Roberts et al., 2013). In particular, a lack of epigenomics study exists in *Lactobacillus*, which is of intense scientific interest. The existing evidence has shown that methylation does not only function in serving as part of a restriction-modification (**RM**) system but also play important roles in chromosome stability, mismatch repair, replication, and gene expression (Low et al., 2001; Marinus and Casadesus, 2009). In the present study, SMRT was used to sequence *L. casei* Zhang and *L. plantarum* P-8 that were originally isolated from dairy products (Wu et al., 2009; Wang et al., 2014), and the newly generated genome data were further compared with the previous published genomes.

Lactobacillus casei Zhang and *L. plantarum* P-8 were obtained from the Collection Centre of Lactic Acid

Received December 22, 2014.
Accepted January 26, 2015.
[1]Corresponding author: hepingdd@vip.sina.com.

ARTICLE IN PRESS

Table 1. General genome information of *Lactobacillus casei* Zhang and *Lactobacillus plantarum* P-8

Strain	Plasmid	Size (bp)	GC content (%)	rRNA	transfer RNA	CDS[1]
L. casei Zhang	1	2,861,969	46.51	15	59	2,807
L. plantarum P-8	7	3,035,719	44.8	16	67	2,952

[1]CDS = coding sequence.

Bacteria at Inner Mongolia Agricultural University of China. The strains were grown in 5.0 mL of de Man, Rogosa, Sharpe broth (Oxoid, Hampshire, UK) at 37°C for 18 to 22 h. The DNA isolation was performed using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI). The integrity of DNA was examined by 0.6% agarose gel and 1.2% lonza gel electrophoresis. The SMRT bell template library preparation was performed as described previously (Eid et al., 2009; Korlach et al., 2010).

The SMRT sequencing was performed on a PacBio RS II instrument (Pacific Biosciences). De novo assemblies were subjected to standard hierarchical genome assembly process (Chin et al., 2013) using just PacBio data from a single, long-insert library, and the consensus was called across reads after assembly polishing. Polymorphic sites were identified with Quiver algorithm available in the Pacific Biosciences SMRT Portal (version 2.3.0). Interpulse durations were measured as previously described (Murray et al., 2012). To identify the common base modifications and analyze methyltransferase motifs, the protocol for modification and motif analysis in the software was selected, which used an in silico kinetic reference and a *t*-test-based kinetic score detection of modified base positions. Visualization of the information from the data analysis was displayed in SMRT View.

Gene prediction was performed by using Glimmer and annotated using BLAST against nonredundant database (provided by the National Center for Biotechnology Information; Altschul et al., 1990; Delcher et al., 2007). The transfer RNA were predicted using tRNAscan-SE (Lowe and Eddy, 1997). The rRNA were predicted using RNAmmer (Lagesen et al., 2007). Function category was predicted with clusters of orthologous genes and Kyoto Encyclopedia of Genes and Genomes databases (Tatusov et al., 1997). Restriction-modification systems were identified by scanning locally for homologs in REBASE (Roberts et al., 2010). The characterization of protein family, domain, and biologically significant sites was carried out using Interproscan (Hunter et al., 2012). The accession numbers for the chromosomes of *L. casei* Zhang and *L. plantarum* P-8 were CP001084 and CP005942, and were CP000935 (plca36), CP005943-CP005948 (LBPp1-LBPp6), and CP010527 (LBPp7) for the plasmids.

Previously, *L. casei* Zhang was sequenced by an Applied Biosystems 3730 DNA analyzer (Grand Island, NY), whereas *L. plantarum* P-8 was analyzed with Roche 454 (Indianapolis, IN) and Illumina sequencing technologies (San Diego, CA). The 2 strains bear 1 and 6 extra chromosomal plasmids, correspondingly (Zhang et al., 2008a, 2010, 2015). Compared with the 2 reference genomes, SMRT sequencing resulted in higher quality, finished genomes for both strains. The de novo assemblies were fully collinear with the reference genomes. The general information of the 2 genomes was described in Table 1.

The hierarchical genome assembly process assembled genome of *L. casei* Zhang resulted in 2 contigs. A total of 137 SNP were identified, which differed from the reference. In *L. plantarum* P-8, 51 SNP and 10 insertions were detected compared with the chromosome reference. The insertion sequences were collapsed repeats uniquely resolved de novo by PacBio, which were missing in the reference. Three of the 7 plasmids were significantly different from the reference sequences achieved from the public DNA databases. The PacBio reads have provided direct evidence that the reference sequences of these 3 plasmids were not correct. Specifically, numerous breaks and areas of poor coverage were found in the shotgun read structure of the plasmid LBPp4. The plasmid LBPp7 was uniquely resolved de novo by PacBio, which was missing in the original reference sequence. Moreover, regional similarity with the reference plasmid of LBPp1 could be identified. Collectively, repeat elements common to both plasmids and chromosomes likely caused the errors in the reference.

In recent years, the next-generation sequencing technologies including the shotgun methods have been commonly used to sequence bacterial genomes. However, the frequency of repeated sequences in the bacterial genomes causes complications and difficulties in the genome assembly process. Our data have clearly indicated that the long reads generated by the SMRT instrument overcome the difficulties brought about by these repeats. Similar successes were reported by Chin et al. (2013), that assembly using only SMRT reads achieved comparable performance relative to other DNA sequencing platforms.

To identify methylated sites within the genomes of *L. casei* Zhang and *L. plantarum* P-8, 2 SMRT cells

Download English Version:

https://daneshyari.com/en/article/10975508

Download Persian Version:

https://daneshyari.com/article/10975508

Daneshyari.com