

# Comparative genomic analysis of 551 *Lactobacillus plantarum* isolates: habitat-specific genomes shaped by frequent recombination

Kangning Li<sup>1</sup>, Shaojing Wang<sup>2</sup>, Wenjun Liu<sup>3</sup>, Lai-Yu Kwok<sup>1</sup>, Bilige Menghe<sup>3</sup>, and Wenyi Zhang<sup>3</sup>

<sup>1</sup>Inner Mongolia Agricultural University

<sup>2</sup>Nankai University

<sup>3</sup>Affiliation not available

May 26, 2020

## Abstract

*Lactobacillus* (L.) *plantarum* is widely distributed in the natural environments and previous studies failed to identify habitat-specific genomic signatures. Thus, comparative genomic analysis together with genome-wide association study of 551 L. *plantarum* genomes were performed to investigate how natural habitats shaped the bacterial genomes. Firstly, the study constructed a phylogenetic tree using a strict core-genome of L. *plantarum*, followed by refining the clades with genome-wide single nucleotide polymorphisms (SNPs). Our data showed that plant-originated isolates dispersed evenly across the phylogenetic tree, supporting a previously reported model of nomadic lifestyle. Isolates from animal and dairy sources showed significant biases in clade distribution, suggesting close association between bacterial genomes and habitats. The animal-originated isolates distributed mainly to clade A and were characterized with more carbon utilization-associated genes. The dairy isolates distributed mainly to subclades B2 and B3, which harbor more nitrogen utilization-associated genes. Both animal and dairy isolation sources were directly linked with the development of human civilization in the aspects of food culture and food industry. Furthermore, based on five population genetic signals, subclade B2 had the highest genetic diversity and SNP distance. Subclade B1 had the highest mutation rate, possibly resulted from recombination events. The animal-originated isolates had lower nucleotide diversity and higher recombination/mutation ratio, which might be the genetic trail remained from the adaptation process. Our data have provided an explanation for the high versatility of L. *plantarum* on the genomic level and showed that the L. *plantarum* species is more genetically diverse than previously expected.

## Hosted file

manuscript.doc available at <https://authorea.com/users/326504/articles/454324-comparative-genomic-analysis-of-551-lactobacillus-plantarum-isolates-habitat-specific-genomes-shaped-by-frequent-recombination>