

**Title: Divergence of microbial community composition and functionality upon serial transfers in novel substrates over 200 generations**

Authors: *Shepherd Nehanda<sup>1,2\*</sup>, Anna Y. Alekseeva<sup>2</sup>, Oscar van Mastrigt<sup>2</sup>, Justin Chileshe<sup>1</sup>, Bas J. Zwaan<sup>2</sup>, Eddy J. Smid<sup>2</sup>, and Sijmen E. Schoustra<sup>2,3</sup>*

*1. National Health Research and Training Institute (Formerly Tropical Diseases Research Centre), Department of Biomedical Sciences, Zambia.*

*2. Laboratory of Genetics and Laboratory of Food Microbiology, Wageningen University and Research, The Netherlands.*

*3. Department of Food Science and Nutrition, The University of Zambia, Zambia.*

*\*Correspondence: Email: [sheperd.nehanda@wur.nl](mailto:sheperd.nehanda@wur.nl).*

This dataset supports chapter 3 of the ongoing PhD project entitled “**Species sorting and community-level functionality in a natural microbial community**”. The data were collected from an adaptive evolution experiment conducted at the Tropical Diseases Research Centre laboratory, Zambia, involving a Zambian traditional dairy fermented milk, known as mabisi. This experiment was aimed at understanding dynamics of bacterial community structural diversity and its functionality when exposed to novel environments over temporal scales spanning about 200 generations. The novel environments were represented by raw cow milk (RCM), F100 infant formula (F100), S26 infant formula (S26), low-fat milk (LFM), and full-cream milk (FCM).

The bacterial community diversity was determined by 16S rDNA amplicon sequencing, along with alpha and beta diversity analysis. The community-level functionality was determined by proxies of volatile organic compound production using gas chromatography mass spectrometry, and pH changes with a pH meter.

This repository is being made public as supplementary data for publication and reference for the general community ecology audience and other scientific readership.

This repository constitutes respective raw datasets, including Excel files for gas chromatography mass spectrophotometry output, pH results, microbiome data (ASV-DNA-sequences.fasta, tree.nwk, ASV table, taxonomy, and metadata), as well as qmd files for R statistical codes.