

**Title: Species sorting shapes the divergence of a natural bacterial community with repeatable functionality during propagation with alternative substrates**

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 2 of the ongoing PhD project entitled “**Species sorting and community-level functionality in a natural microbial community**”, involving mabisi, a Zambian traditional dairy fermented milk as a model. The data were collected from a field based evolution experiment in Zambia, which aimed to explore species sorting process and its consequences on the community-level functionality *in situ*. 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 profiling volatile organic compounds using gas chromatography mass spectrometry, pH changes with a pH meter, and consistency with an Adam’s consistometer.

This repository is being made publicly available as supplementary data to support the associated publication and as a reference to the readership in microbial community ecology, food biotechnology, and other scientific audiences.

This repository constitutes respective raw datasets, including excel files for output from high-pressure liquid chromatography, gas chromatography mass spectrophotometry, pH and consistency results, microbiome data (zipped files from the Illumina sequencing machine, ASV-DNA-sequences.fasta, tree.nwk, ASV table, taxonomy, and metadata), as well as qmd files for R statistical codes.