P137

Monitoring strain diversity in metagenomics data using meta-MLST SNP profiling

V.C.L. de Jager[1,2], M.R. van der Sijde[3], R. Hagelaar[4], L.A. Hazelwood[3], O. Kutahya[1], M. Kleerebezem[1], E.J. Smid[1], R.J. Siezen[4], S.A.F.T. van Hijum[3]

[1]Wageningen University, Dept. of Microbiology, Wageningen, [2]the Netherlands Bioinformatics Centre, [3]NIZO food research, Ede, [4]Center for Molecular and Biomolecular Informatics, Nijmegen

Introduction: Microbial consortia, or complex (undefined) mixtures of microbes, are ubiquitous in nature. They are found everywhere ranging from soil to gut and from biofilms to industrial fermentations.

Monitoring bacterial diversity in consortia from metagenomics sequence data, is typically done using e.g., 16S pyrosequencing techniques. These techniques mostly allow describing a bacterial consortium up to the genus level.

The assumption in many metagenomics studies and the subject of many published reviews is that knowing what microbial taxa are present also indicates their functionality in the community. Based on the limited knowledge about bacterial strains and their gene content, and thus functionality, the extrapolation of knowing what species is present to their functionality is therefore highly challengeable. In addition, in many environments, e.g. from acid mine drainage, human-controlled aquatic environments and dairy starter cultures, coexistence of multiple closely related strains is observed, of which the diversity is not only determined by their unique gene content, but also by plasmid content and phage sensitivity. Our understanding of a population's complexity will therefore largely depend on the ability to differentiate between genetically highly similar individuals.

Method: We describe a method that allows following the naturally occurring bacterial diversity at the strain level. It involves selecting 'core' genes in a set of isolated representative strains that are expected to be present in all bacterial strains present in the metagenome. Next, (combinations of) single nucleotide polymorphisms (SNPs) are determined that allow distinguishing (groups of) strains. These SNPs are subsequently used to categorize reads or contigs obtained from next generation sequence analysis of metagenomic samples into strains, similar to multi locus sequence typing (MLST).

Results: We apply this technique to follow strain level diversity of *Lactococcus lactis* in multiple-timepoint metagenomics data obtained during the cheese making process. The resulting strain types are mapped on metabolic pathways.

Conclusions: We show that 1) following highly similar individual strains directly from metagenome sequence data is feasible and 2) mapping of identified strain types on metabolic pathways gains insight in the metabolic potential of individual isolated strains vs. the metagenome.