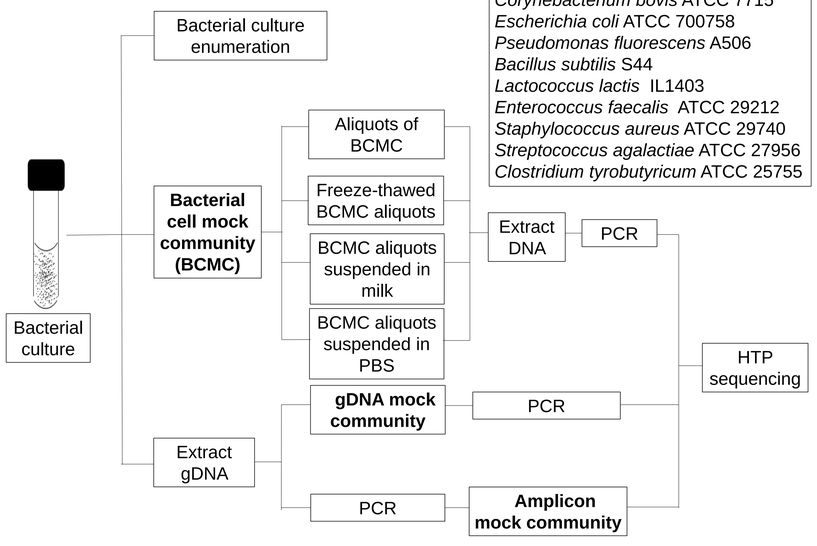
Accurate monitoring of living and total bacterial populations in milk for improved microbial management

Zhengyao Xue, Mary Kable and Maria L. Marco
Department of Food Science & Technology, University of California, Davis

Background and objectives

Milk contains a broad diversity of bacteria from farm and processing environments. These bacteria are important for the quality and safety of dairy products. Currently employed methods for microbial analyses of milk typically target only a limited number of bacterial species. Therefore, we are developing and applying methods for bacterial diversity measurements using culture-independent, high-throughput (HTP) 16S rRNA gene sequencing. Herein, we describe the application of a bacterial cell and DNA mock community comprised of nine strains of dairy-associated bacterial species to assess protocols for the detection of bacterial populations in milk.

Methods



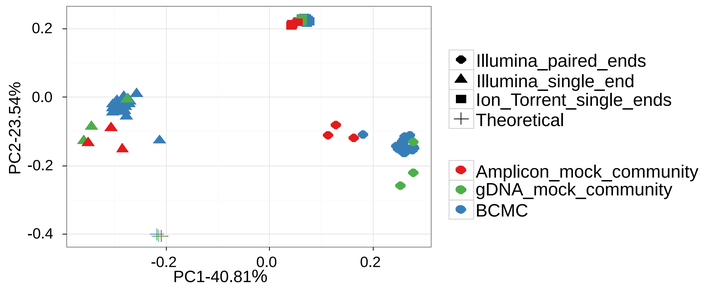
Bacterial strains used in this project
Corynebacterium bovis ATCC 7715
Escherichia coli ATCC 700758
Pseudomonas fluorescens A506
Bacillus subtilis S44
Lactococcus lactis IL1403
Enterococcus faecalis ATCC 29212
Staphylococcus aureus ATCC 29740
Streptococcus agalactiae ATCC 27956
Clostridium tyrobutyricum ATCC 25755

BCMC, gDNA and amplicon mock communities were prepared in triplicate. DNA was extracted with two commercial kits (MagMAX Total Nucleic Acid Isolation Kit and PowerFood Microbial DNA Isolation Kit) from the BCMC. The Illumina Miseq 250 paired-end and Ion Torrent PGM 400 single-end methods **(Ion Torrent single end)** were used for HTP DNA sequencing. Sequence reads from Illumina Miseq were either joined after quality trimming (Phred score >30) with 100bp overlap and 1% error (**Illumina paired ends**) or not joined (**Illumina single end**). Additionally, an *in silico* mock community was created with 16S rRNA gene V4 amplicon sequences proportional to the taxa relative abundances in each mock community (**Theoretical**). Sequence analysis was performed in QIIME 1.9.1 with 97% similarity for Operational Taxonomic Units (OTUs) clustering using the Greengenes_13_8 database. Low abundance OTUs (singletons and <0.05%), chloroplast and Archaeal OTUs were dropped from the final OTU table.

Acknowledgements

This project was supported with funds from the California Dairy Research Foundation.

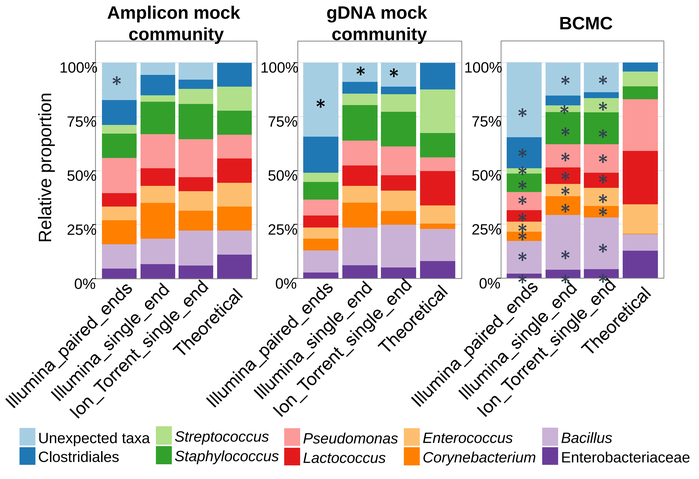
HTP 16S rRNA DNA sequencing methods impact assessments of bacterial diversity



Principal coordinate analysis (PCoA) of the Bray-Curtis distances between the mock communities. The OTU table was normalized by cumulative sum scaling (CSS).

- Sequencing method has the greatest impact on assessments of bacterial diversity. (**79.81%** of the total variation (adonis, p<0.001))
- DNA extraction method and the presence of the milk matrix conferred minor effects.

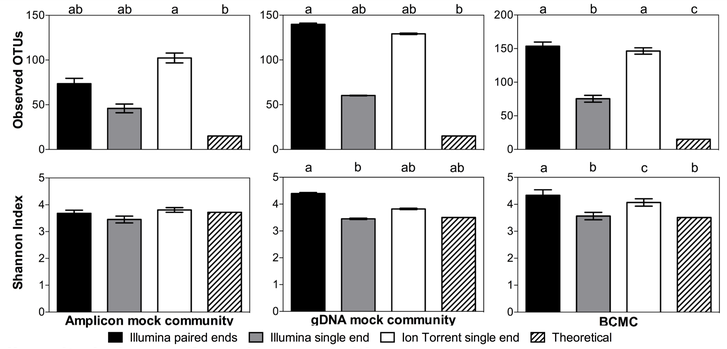
Single end sequencing methods have fewer erroneous taxonomic assignments



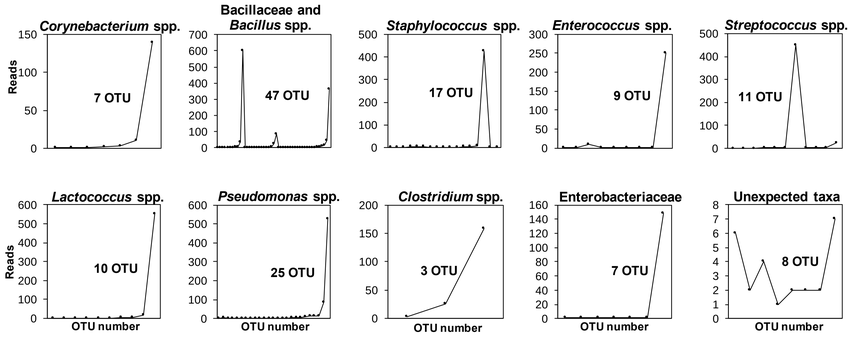
Relative proportion of taxa in the mock communities. Expected taxa (9 bacterial species) are labelled with the corresponding taxonomic level from sequencing results. Values shown are the average of 24 replicates of BCMC and 3 replicates of gDNA and amplicon mock communities. CSS normalized OTU tables were used for making the taxonomy summary. Significant differences in taxa abundances between theoretical and sequencing results are marked with asterisks (Mann-Whitney U test, p<0.05).

- The Illumina paired ends method has the most sequencing errors.
- According to the gDNA mock community, all sequencing methods resulted in unexpected taxa.
- The distribution of bacterial taxa in the BCMC observed was significantly different from the theoretical values for all sequencing methods .

Sequencing method significantly inflates the alpha diversity of mock communities



Alpha diversity indexes of mock communities. Values shown are the avg ± ste of 24 replicates of BCMC and 3 replicates of gDNA and amplicon mock communities. Both observed Shannon index and observed OTUs are normalized for 4,300 sequences per sample. Significant differences (p<0.05) are indicated by the presence of different lowercase letters above each boxplot. Multivariate analysis was performed with Kruskal-Wallis test followed by Nemenyi-test pairwise comparisons.



OTU clustering of data collected from the Ion Torrent single end sequencing. OTU table was normalized for 4,300 sequences per sample. Results for a single replicate are shown.

- Increases in the number of observed OTUs are partially caused by multiple OTU assignments to individual (expected) taxa.
- The unexpected OTUs are closely related members of the mock community. The "unexpected taxa" included: *Anaerobacillus* (1 OTU), *Bacillus cereus* (2 OTUs), Planococcaceae (4 OTUs), *Enterococcus casseliflavus* (1 OTU).

Conclusions

- Mock communities are ESSENTIAL tools for validating 16S rRNA gene sequencing methods, yet are rarely applied.
- Errors were mainly introduced by HTP DNA sequencing platforms and bioinformatics analyses (joining paired ends).
- The number of observed OTUs is not a robust index for comparing the intra-sample diversity due to the multiple OTU assignments to one taxon and erroneous OTU assignments to closely related organisms.
- Single end DNA sequencing in combination with the MagMAX Nucleic acid DNA extraction kit improved the accuracy of diversity measurements.

Impact to the dairy industry

Using a stream-lined, automated, HTP 16S rRNA gene sequencing workflow to identify bacterial species in milk and dairy processing environments will enable improved microbial management and result in consistent products with optimal sensory profiles and nutritive benefits.