Microbiome Profiling at the Strain Level Using rRNA Amplicons

Bo-young Hong1, Dawn Gratato2, C. Clark2, Thomas Jarvie2, George M. Weinstock1, Mark Driscoll1

1 The Jackson Laboratory for Genomic Medicine, Farmington, CT, 2 Shoreline Biome, Farmington, CT

Background: Strain level microbiome profiling is needed for a full understanding of how microbial communities influence human health. Traditionally, microbiome profiling of rRNA gene amplicons is a well-understood method that is rapid and inexpensive, but standard 16S rRNA gene methods generally cannot differentiate closely related strains. Whole genome shotgun microbiome profiling is considered a higher-resolution alternative, but with decreased throughput and significantly increased sequencing costs and analysis burden. With both methods there are also challenges with microbial lysis, DNA preparation, and taxonomic analysis. Specialized microbiome-focused protocols were developed to achieve strain-level taxonomic differentiation using a rapid, high throughput rRNA gene assay. The protocol integrates lysis and DNA preparation improvements with a unique high information content amplicon and associated novel database to enable taxonomic differentiation of closely related microbial strains.

Methods: Comprehensive Lysis: A novel, rapid, microbiome lysis system opens bacteria without bead beating to obtain DNA from lysis-resistant microbes without damaging DNA. Fecal samples, commonly used to collect the microbiome, were prepared using commercially available DNA preparation kits from multiple manufacturers, and microbiome profiles were compared to the novel method after sequencing.

Results: 16S rRNA Image from: http://rna.ucsc.edu/rnacenter/images/figs/ecoli_16s.pdf

PCR Design Targeting Primer Coverage of rRNA PCR Site Variants: Ribosomal RNA 2-dimensional folding constraints were analyzed for common PCR primer sites across the Riken, Silva, and Athena databases. Sequences at the 271 primer site in the V1 region were extracted from all organisms in each database using a custom Python program. The sequences were compared to the E. coli 271 folded structure based on the E. coli image at the UCSC Center for Molecular Biology of RNA at (http://rna.ucsc.edu/rnacenter/bosomi_images.html).

A novel lysis and DNA purification protocol was developed to provide a rapid, sensitive, and effective method to collect DNA from heavily lysis-resistant samples. The protocol was then compared to an established method.

Conclusions: Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Ext amplicon reads from four high-E. coli human fecal samples were mapped to the Athena database of over 4,000 16S rRNA gene sequences. The database included 137 different E. coli genomes in the Athena database, each containing 7 16S-23S regions. Sample 4CancerEXT showed high levels of amplicon reads from the strain E. coli LM026E. Other samples contained reads from different strains.

Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Within each sample, reads mapped to Athena database that matched uniquely to different 16S-23S regions inside a single sample. As an example, reads from Sample 4Cancer EXT from above showed aligned to regions 1, 5, and 7 of the E. coli LM026E genome, demonstrating multiple hits to different regions of the E. coli LM026E genome. The circular E. coli LM026E genome is shown with approximate locations of the individual 16S-ITS-23S regions.

Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

137 different complete E. coli genomes were compared in silico. About 20 strains contained one or more uniquely identifiable V4 region sequences whereas overall strain contained one or more unique sequences in the EXT amplicon. Variants in E. coli 16S region is found mostly in V1, V3 and V4 regions. The long read amplicon/Athena database combination can be used to differentiate closely related E. coli strains (right).

Conclusions: Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Conclusions: Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Conclusions: Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Conclusions: Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Strain-level Taxonomic Differentiation Depends on ITS Region of EXT Amplicon:

Within each sample, reads mapped to Athena database that matched uniquely to different 16S-23S regions inside a single sample. As an example, reads from Sample 4Cancer EXT from above showed aligned to regions 1, 5, and 7 of the E. coli LM026E genome, demonstrating multiple hits to different regions of the E. coli LM026E genome. The circular E. coli LM026E genome is shown with approximate locations of the individual 16S-ITS-23S regions.

Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

1719 different complete E. coli genomes were compared in silico. About 20 strains contained one or more uniquely identifiable V4 region sequences whereas overall strain contained one or more unique sequences in the EXT amplicon. Variants in E. coli 16S region is found mostly in V1, V3 and V4 regions. The long read amplicon/Athena database combination can be used to differentiate closely related E. coli strains (right).

Conclusions: Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.

Conclusions: Long Amplicon Strain-level Microbial Identification in Fecal Samples: Athena Database Assigns Strain-Level Taxonomy.