Session Title: EEB03 - Microbial Biodiversity and Systematics II
Date and Time: 6/23/2019 10:30:00 AM - 6/23/2019 4:00:00 PM
Location: Exhibit and Poster Hall
POSTER BOARD NUMBER: SUNDAY - EEB-523

Microbiome Profiling at the Strain Level Using rRNA Amplicons

Bo-young Hong[1], Dawn Gratalo[2], C. Clark[2], Thomas Jarvie[2], George M. Weinstock[1], Mark Driscoll[2] [1] The Jackson Laboratory for Genomic Medicine, Farmington, CT, [2] Shoreline Biome, Farmington, CT

Shoreline Biome
Mark.Driscoll@shorelinebiome.com
(203) 627-9749

Background:

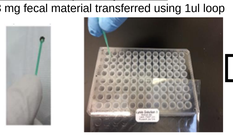
Strain level microbiome profiling is needed for a full understanding of how microbial communities influence human health. Microbiome profiling of rRNA gene amplicons is a well-understood method that is rapid and inexpensive, but standard 16S rRNA gene methods generally cannot differentiate closely related 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-specific lysis opens bacteria without bead beating to obtain DNA from lysis-resistant microbes without damaging DNA. Fecal samples, custom and commercial mock microbiomes were prepared using commercially available DNA preparation kits from multiple manufacturers, and microbiome profiles were compared to the novel method after sequencing.

Novel Lysis and DNA Purification

Transfer Sample to Lysis Plate
3 mg fecal material transferred using 1ul loop



Transfer 96 Fecal Samples
~15 mins

Parallel Rapid 96-sample Protocol
Add KOH to Plate- 2 min
Heat Plate- 5 min
Spin Plate-2 min
Transfer supernatant- 2 min
Wash DNA magbeads- 5 min
Elute- 5 mins

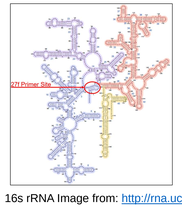
Purified DNA in 96-well format
~20 mins

rRNA PCR and Sample Barcoding
Transfer 10ul DNA to PCR Plate- 2 mins
Add 10ul of 2x PCR Premix- 2 mins
Cover plate and place on thermocycler

PCR ready in 96-well format
~4 mins

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 27f 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* 2-D folded structure based on the *E. coli* image at the UCSC Center for Molecular Biology of RNA at (http://rna.ucsc.edu/rnacenter/ribosome_images.html). Sequence variants were mapped to the base paired structure to determine whether variation altered 16S rRNA gene- structures. PCR primer pools that cover all known bacterial primer site sequence variants were created for rRNA amplicon sequencing.

16S rRNA Structure



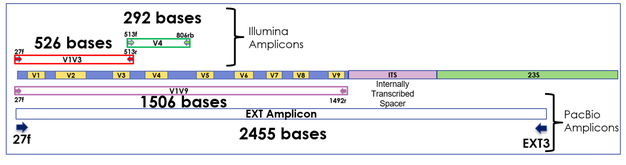
Primer Number	Matching Regions in Database			27f Primer Sequence
	Athena	Riken	SILVA SSU	
1	26221	8339	15659	AGAGTTTGATCCTGGCTCAG
2	15786	3634	9013	AGAGTTTGATCATGGCTCAG
3	254	146	244	AGAGTTTGATTCTGGCTCAG
4	277	136	62	AGGGTTCGATTCTGGCTCAG
5	292	82	28	AGAATTTGATCTTGGTTCAG
6	137	73	97	AGGGTTTGATCCTGGCTCAG
7	80	35	77	AGAGTTCGATCCTGGCTCAG
8	200	31	78	AGAGTTTGATCTTGGCTCAG
9	23	14	59	AGAGTTTGATCCTAGCTCAG
10	15	31	26	AGAGTTTGATCCTGGCTTAG
11	4	5	10	AGAGTTTGATCCTTGCTCAG

Degenerate Consensus
AGRRTTYGATYHTDGYTYAG

Bacterial primer 27f variants identified in the Athena 16S-23S, Riken and SILVA SSU database. A custom Python program identified the closest match to the consensus 'AGAGTTTGATCCTGGCTCAG' at or near position 27 in the 16S genes of bacterial sequences in the each database. Primer sequences were assigned a number in column 1. The number of database entries with matches for each sequence at position 27 are listed in columns 2, 3 and 4. Red text in column 5 signifies a change from Primer 1 sequence. The degenerate consensus incorporates 27f variation from all 11 primers, with invariant bases in black.

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

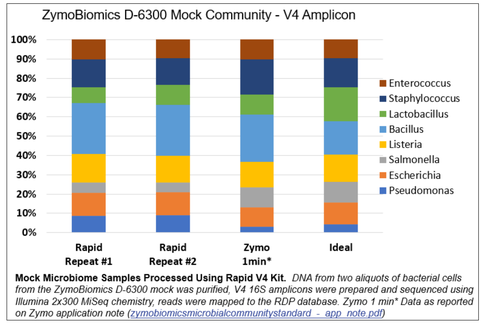
Long Amplicon Strain-level Microbial Identification: DNA from twenty fecal microbiome samples was profiled using a ~2500 base amplicon (EXT) that spans the entire 16S gene and extends into the 23S gene. EXT amplicon ccs reads sequenced using PacBio Sequel were mapped to a novel database (Athena) created from contiguous 16S-23S gene sequences from complete microbial genomic assemblies. Additionally, reference genome sequences of EXT, V1V9, V1V3 and V4 amplicons from 127 strains of *E. coli* were compared *in silico* using a custom Python program designed to identify sequence differences between strains, and strains with unique sequences were identified for each amplicon.



Amplicons Used in this Study. The center of the figure shows the full 16S-23S gene regions on either side of the highly variable ITS region. V4, V1V3, V1V9 and EXT amplicons are shown, with amplicon lengths.

Results:

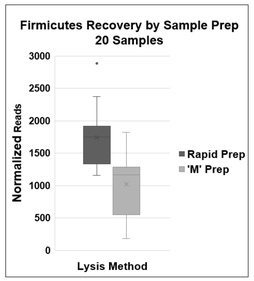
Bead Beating Not Required. Accurate microbiome profiling methods must obtain DNA from all microbes. Rapid method was used to prepare a commercially available ZymoBiomics 8-bacteria mock microbiome, V4 amplicons were prepared, and microbial profiles were compared to manufacturer's results (right)

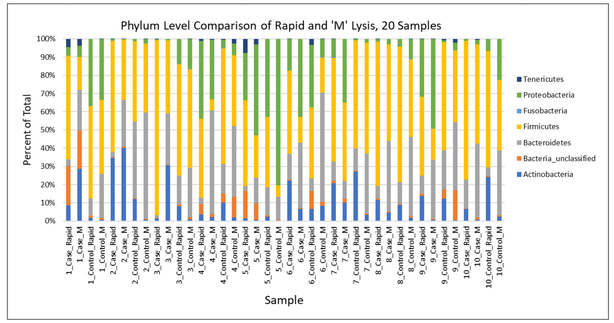


Mock Microbiome Samples Processed Using Rapid V4 Kit. *DNA from two aliquots of bacterial cells from the ZymoBiomics D-6300 mock was purified, V4 16S amplicons were prepared and sequenced using Illumina 2x300 MiSeq chemistry, reads were mapped to the RDP database. Zymo 1 min* Data as reported on Zymo application note (zymobiomicsmicrobialcommunitystandard_-_app_note.pdf)*

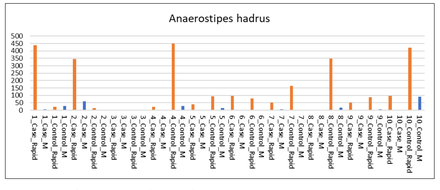
Comprehensive Lysis:. Twenty human fecal samples (below) were processed with two lysis and DNA purification methods, the Rapid method and commercial method "M".

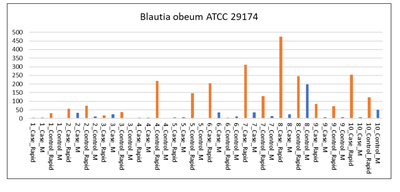
Rapid Method Increased Firmicutes Representation in Fecal Samples: Comparison of Rapid lysis with commercial method 'M' revealed that representation of Gram-positive Firmicutes was increased by almost 30% in fecal samples





Firmicutes Appear in Rapid Lysis Samples: EXT amplicon analysis identified specific Firmicutes strains with improved representation. Two example strains that are common to most of the 20 fecal samples prepared using the EXT 16S-23S amplicon are shown. The appearance of the individual Firmiciutes strains is consistent with the phylum level results showing a 30% increase in Firmicutes.





PCR Design Targeting Primer Coverage of rRNA PCR Site Variants: Variation within conserved rRNA gene sequences was found to maintain base pairing in proximal stem-loop structures, or maintain structural base pairing over long distances. For example, base changes in the 27f stem were complementary, with a G/C base pair exchanged for an A/T base change. Changes to the loop in the 27f V1 sequence were complemented by base pairs changes in the V4 region 600 bases away that were required to maintain rRNA structure. The structural constraints are consistent with the limited 27f primer site base variation found across the thousands of organisms and wide variety of phyla in the databases. New primer design guidelines were developed based on structural conservation to guide the creation of rRNA PCR primer pools containing fewer than 20 primers that are capable of targeting all known bacterial primer site sequence variants.

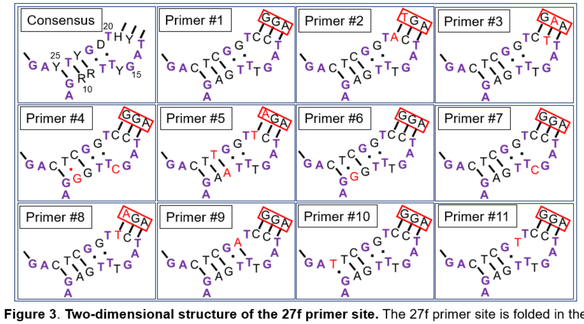
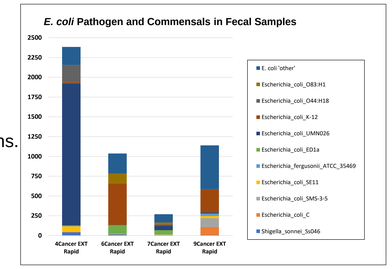


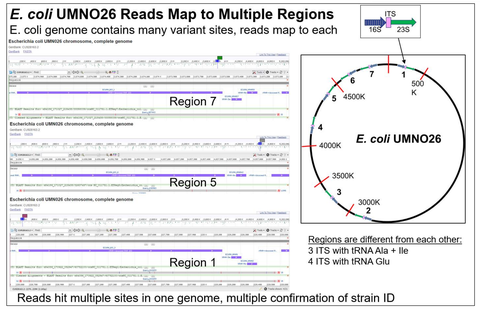
Figure 3. Two-dimensional structure of the 27f primer site. The 27f primer site is folded in the rRNA stem-loop structure, where conserved (invariant) bases are in bold purple text, with a 'dash' indicating consensus base pairing and a 'dot' representing the G•U wobble base pairing common in rRNA. Red bases represent bases that differ from Primer #1 sequence. Consensus sequence with degenerate base sequence is shown in the top left panel. Primers with sequence variations in the stem are shown, except for Primer #5, which has changes in the stem and loop, and details corresponding base pairing to a sequence in the V4 region ~900 bases away (red box). The base paired structure and consensus numbering is based on the *E. coli* image at the UCSC Center for Molecular Biology of RNA at (http://rna.ucsc.edu/rnacenter/ribosome_images.html).

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 40,000 16S-23S gene sequences (right). There are 137 individual *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 UMNO26*. Other samples contained reads from different strains.

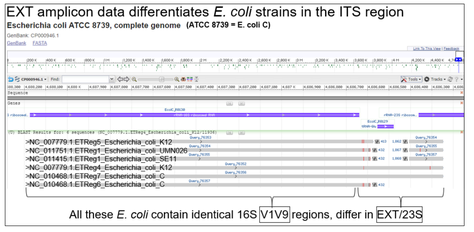




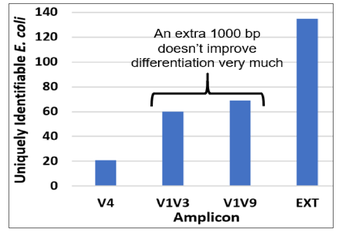
Within each sample, reads mapped to Athena database were identified that mapped uniquely to different 16S-23S regions inside a single strain. As an example, reads from Sample 4Cancer EXT' from above are shown aligned to regions 1,5, and 7 of the *E. coli* UMNO26 genome, demonstrating multiple hits to different regions of the *UMNO26* genome The circular *E. coli UMNO26* genome is shown with approximate locations of the individual 16S-ITS-23S regions.

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

Visualization of differences between *E. coli* strain EXT amplicons: EXT amplicon reads from *E. coli C*, *K-12*, *SE11*, and *UMN026* were all mapped to the *E. coli C* genome. All have identical V1V9 regions, but can be differentiated by sequence in the Internally Transcribed Spacer region (right).



137 different complete *E. coli* genomes were compared *in silico*. About 20 strains contained one or more uniquely identifiable V4 regions, whereas every strain contained one or more unique sequences in the EXT amplicon. Variation in *E. coli* 16S region is found mostly in V1-V3, explaining why full-length 16S does not add many unique regions. The long read amplicon/Athena database combination can be used to differentiate closely related *E. coli* strains (right).



Conclusions:

Strain level taxonomic differentiation of closely related microbes within human fecal microbiomes is relatively straightforward using a modified rRNA based assay. The combination of robust lysis, comprehensive PCR primer pools, ~2500 base EXT amplicon product, and long read Athena database enables differentiation of closely related strains within and across samples using a simple, rapid, cost effective, well-understood rRNA gene targeting approach. *E. coli* amplicon reads from fecal samples were shown to map uniquely to individual genomic regions within the Athena database, and *in silico* experiments using the EXT amplicon sequences from *E. coli* demonstrated taxonomic differentiation of closely related strains.

Citations:
27f sequence: Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., & Olsen, G. J. (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Applied and environmental microbiology, 74(8), 2461-70.
SILVA SSU Database: Christian Quast, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, Frank Oliver Glöckner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, Nucleic Acids Research, Volume 41, Issue D1, 1 January 2013, Pages D590–D596, https://doi.org/10.1093/nar/gks1219
Riken 16S Database:http://metasystems.riken.jp/grd/
16S Structure:*E.coli 16S rRNA Secondary Structure* http://rna.ucsc.edu/rnacenter/images/figs/ecoli_16s.pdf