Targeted Sequencing of 16s rRNA Gene to understand the diversity and composition of the gut microbiome

Rajesh Kumar Gottimukkala, Asha Kamat, Alice Zheng, Karen Clyde, Fiona Hyland, Janice Au Young, Thermo Fisher Scientific,South San Francisco,CA,USA

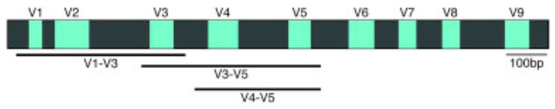
ABSTRACT

Recent studies in humans and experiments in mouse models demonstrated the key role of the gut microbiota in modulating the tumor response to check point blockade immunotherapy. One study showed an association between negative outcome using CTLA-4 blockade therapy and the absence of a specific gut microbiome. So, the gut microbiota has emerged as a promising biomarker to assess the efficacy of immune-modulatory drugs. Next generation sequencing of the 16S rRNA Gene is widely used as standard for understanding the composition of the gut microbiome. The Ion AmpliSeq Pan-Bacterial Community panel that contains 24 primer pairs targeting the 16S rRNA gene provides a cost-effective approach to identify the bacterial species present in the sample. The panel also includes gene-specific markers targeting 21 different species and drug-resistance markers.

INTRODUCTION

Amplicon Type	Number of Targets Detected	Pool	Number of Amplicons
Antibiotic resistance genes	364 genes	1	716
Species Identification	21 species	1	269
Universal 16S primer pairs	99.9% of ~400,000 16S sequences in public Greengenes database	2	24

The 16S rRNA Gene is highly conserved region between different bacteria and is widely used as standard for detection and identification. Due to highly homologous nature of these 16S sequences as available in public databases like GreenGenes, it is challenging to correctly identify different bacteria at the Genus/Species level using short sequencing reads.



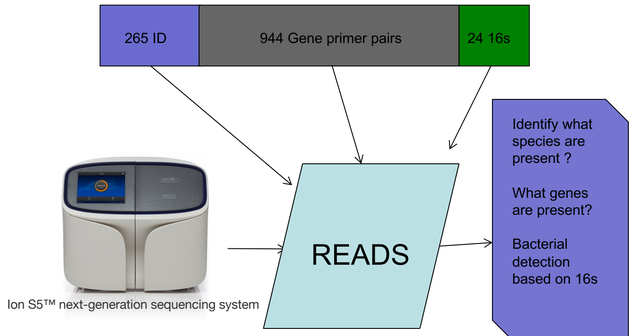
16S rRNA gene structure illustrating variable (blue, V1–V9) and conserved (dark gray) regions. Common regions (amplicons) for next-generation sequencing-based microbial community profiling are depicted below.

Table 1. Overview of internal curated version of public Green Genes 16s database

Summary	Internal curated version of Green Genes database
Number of Entries	146822
Number of different species	13217
Number of different Genus	2241
Number of different Families	346

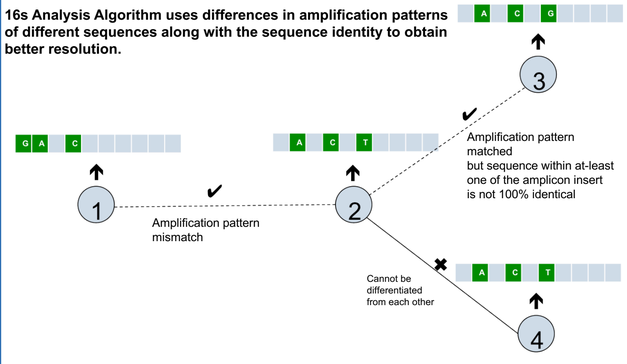
Microbe Identification Pipeline

We have developed a new algorithm that can correctly identify almost all the bacteria species at Genus/Species level. In this method, we align the reads against all the sequences in the database and build a coverage pattern per sequence. It involves separating reads based on the originating primer pairs and computing coverage separately. This part is particularly challenging for Ion Ampliseq reads as the reads do not contain primer sequences, but we are able to solve this using a computational approach. By matching the observed pattern per sequence with an expected pattern that is pre-computed we are able to significantly improve the Species/Genus level resolution.



Overview of the Analysis

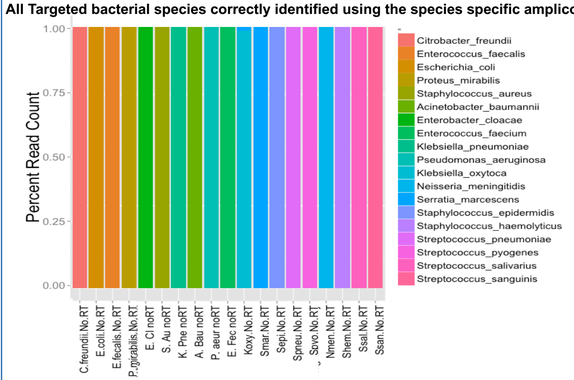
16s Analysis Algorithm uses differences in amplification patterns of different sequences along with the sequence identity to obtain better resolution.



Given a set of primer pairs, any two 16s sequences can be identified from each other if any of these two conditions is true i) the two sequences have different amplification patterns i.e. different sets of primer pairs amplify these two sequences as determined by in-silico e-PCR approach. Ii) the sequences are different within at-least one of the amplicon inserts. In the example in the figure, sequences 2 and 4 cannot be differentiated from each other whereas sequences 1 and 3 can be differentiated from all the other sequences.

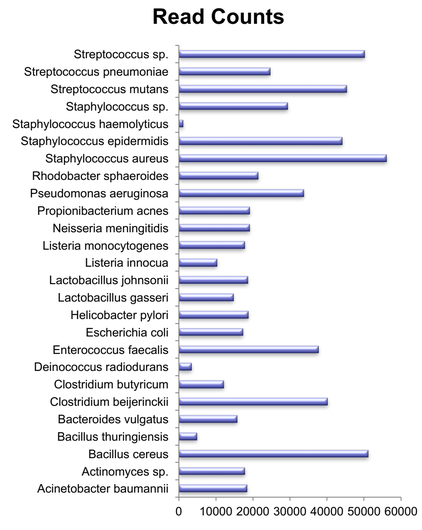
Results

All Targeted bacterial species correctly identified using the species specific amplicons



100% identification specificity achieved with the gene specific targets using the custom reference file created in the analysis pipeline. In the plot, different samples are shown on the x-axis and percent read count of different bacterial species are shown on the y-axis.

Results using 20 Strain Even Mix Genomic Material ATCC® MSA-1002™

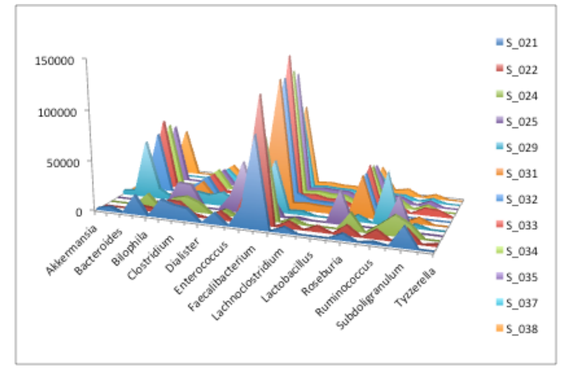


We sequenced a metagenomics mock community sample comprising of 20 different strains and identified all the 20 species including few organisms relevant to cancer microbiome studies like H.pylori, E.Faecalis, B.vulgatus etc.

We identified all the 20 different strains used at Genus level with 100% Sensitivity and Specificity.

At species level, we have 100% sensitivity but three false positive species.

Results from DNA Sequencing of 12 fecal samples



Using the Pan Bacterial assay containing only 24 primer pairs targeting 16S rRNA with amplicon lengths less than 150bp, we sequenced DNA from 12 fecal samples with the assay using Ion GeneStudio S5 System and detected the 25 frequently observed Genera across all the samples including Bifidobacterium, Lactobacillus, Clostridium, Ruminococcus and Bacteroides etc.

We did an in-silico analysis using the primers in the assay and demonstrated that using the assay we can identify the few bacterial microbes in Gut microbiome (few shown below) that are known to be associated with the tumor response.

Species :
Bacteroidaceae Bacteroides thetaiotaomicron
Enterobacteriaceae Escherichia coli
Ruminococcaceae Anaerotruncus colihominis

Genus :
Bacteroidaceae Bacteroides
Ruminococcaceae Faecalibacterium

Family :
Lactobacillaceae
Clostridiales
Ruminococcaceae

CONCLUSIONS

The Ion AmpliSeq Pan-Bacterial Community panel with the described Bioinformatics pipeline will enable potential future usage of 16s rRNA sequencing to assess the Gut microbiome as a potential biomarker for immunotherapy

References

1. Jay-Hyun Jo, Elizabeth A. Kennedy, Heidi H. Kong, Research Techniques Made Simple: Bacterial 16S Ribosomal RNA Gene Sequencing in Cutaneous Research,
2. V. Gopalakrishnan *et al*., *Science* 10.1126/science.aan4236 (2017).