Amplicon Based 16S Ribosomal RNA Sequencing and Genus Identification

*Jan Risinger, *Lindsey Renken, +Josh Hill,
+Noushin Ghaffari, PhD, +Richard P. Metz, PhD,
+Charles D. Johnson, PhD, *Masoud M. Toloue, PhD
*Bioo Scientific
+AgriLife Genomics and Bioinformatics Service, Texas A&M University

ABSTRACT

Next generation sequencing analysis of 16S ribosomal RNA (rRNA) is commonly used to identify bacterial species and perform taxonomic studies. Bacterial 16S rRNA genes contain 9 hyper-variable regions with considerable sequence diversity among different bacterial species and can be used for species id. Rapid determination of highly complex bacterial populations through targeted amplification can provide an accurate gauge of diversity at taxonomic hierarchies as low as the genus level. A single 16S rRNA hyper-variable domain does not have enough sequence diversity to distinguish genera. With increased read lengths of Illumina MiSeq chemistry, Bioo Scientific has expanded the common analysis of the fourth hyper-variable domain (V4) of prokaryotic 16S rRNA to V1, V2 and V3 regions simultaneously. Optimized preparation through a streamlined standardized procedure allows for high-quality, reproducible libraries. This optimization can be applied to different windows of 16S rRNA as well as other relevant prokaryotic taxonomic markers.

METHODS

DNA Isolation and Microbiome Enrichment

DNA was isolated from human saliva using the QIAGEN DNeasy Blood & Tissue kit with minor modifications (1). Quality and quantity of DNA was assessed by spectrophotometry. DNA extracted from saliva was enriched for microbial DNA, and DNA quantity was determined by fluorometer.

16S V1-V3 Library Preparation

20 ng of microbial enriched DNA was used as starting material for a NEXTflex™ 16S V1-V3 Amplicon-Seq library prep. Targeted PCR amplification of the 16S V1-V3 region was performed using the universal primers contained in the kit, which contain library-specific overhangs and are complementary to the conserved domains flanking the hyper-variable regions of interest. After AMPure XP bead cleanup, a subsequent PCR was performed with an indexing set of primers containing Illumina flow cell binding sites, sequencing primer complementary sequences compatible with paired-end sequencing, and indexing barcodes for high-throughput multiplexing of up to 384 unique libraries (Figure 2).

Sequencing and Data Analysis

The pair-end reads from seven samples were used in our analysis. We trimmed the adapters using Cutadapt software. The trimmed reads were analyzed using Metagenomics Rapid Annotation using Subsytems Technology (MG-RAST). The input data was preprocessed by removing artificial replicate sequences and any host specific sequences (H. Sapiens). Reads that has five bases with quality score less than twenty also were filtered out. The quality control of the MG-RAST pipeline identified at least 86% of each libraries data as ribosomal RNA. The taxonomic ranks (domain, phylum, class, order, family, genus), taxon abundance, and GC distribution for all samples were reported.

INTRODUCTION

Before the development of high-throughput methods to identify and characterize microbial populations, our understanding of the role microbes play in environmental, agricultural, and health-related settings was limited. Determining the complexity of species present in a sample can be achieved by sequencing a genomic region, conserved in all species, that contains evolutionarily divergent sequences that allow identification of unique taxa.

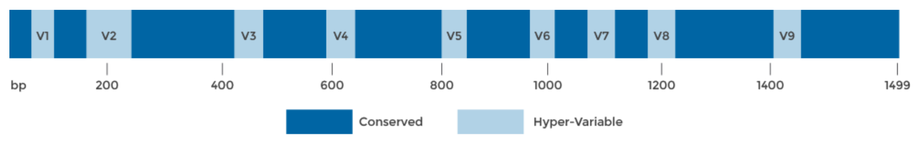

Figure 1. Schematic representing conserved and hyper-variable regions of the 16S rRNA gene

A commonly used phylogenetic marker in metagenomics is the 16S ribosomal RNA (rRNA) gene. This ubiquitous locus is comprised of highly conserved regions flanking nine hyper-variable regions, referred to as V1-V9 (Figure 1). Here we demonstrate the utility of the NEXTflex™ 16S V1-V3 Amplicon-Seq Kit combined with the longer read chemistry of Illumina MiSeq (2x300) for enabling accurate identification of genera present in highly complex microbial communities across a vast number of samples.

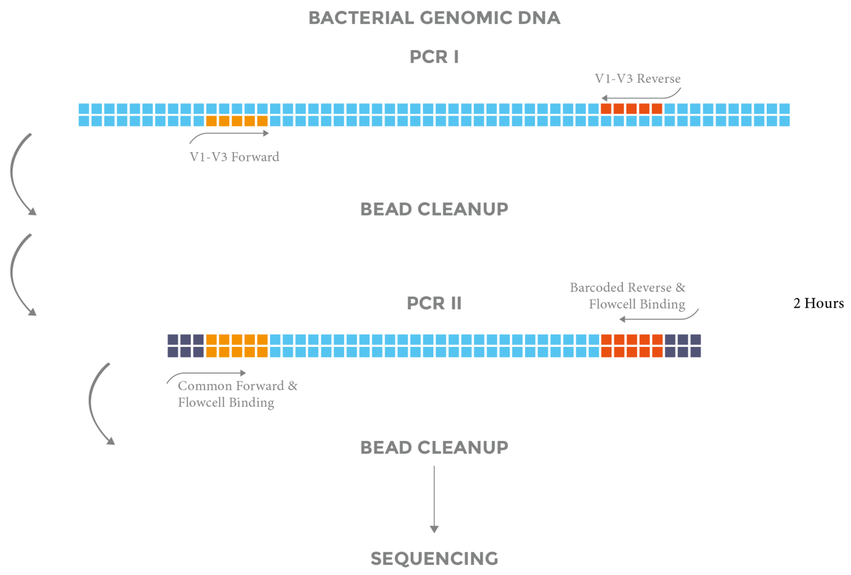

Figure 2. NEXTflex 16S V1-V3 Amplicon-Seq library preparation workflow

RESULTS

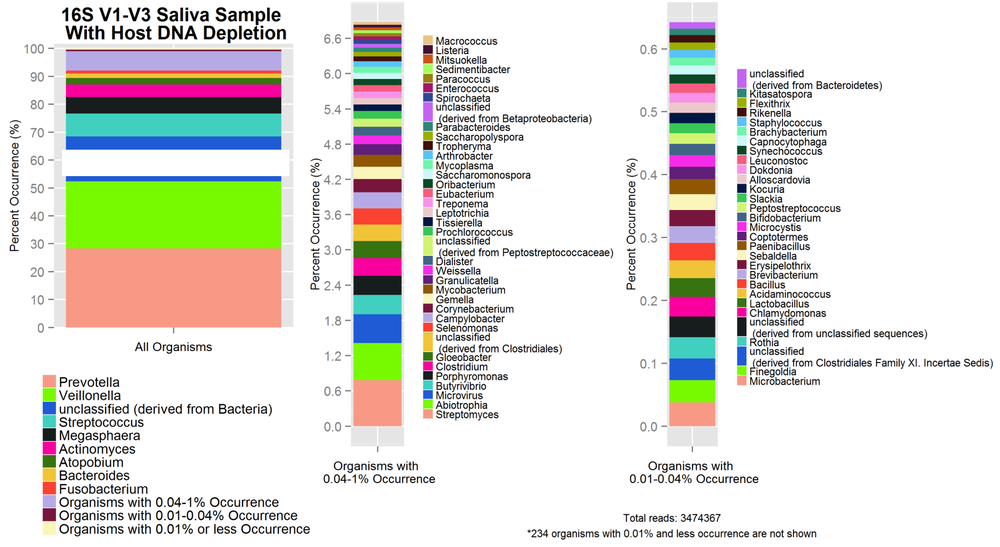

Figure 3. NEXTflex 16S V1-V3 Amplicon-Seq analysis of organisms present in human saliva sample

We explored the microbial community composition in human saliva using the NEXTflex 16S V1-V3 Amplicon-Seq Kit. High proportions of the genera *Veillonella* and *Streptococcus* were identified (Figure 3). *Veillonella* requires the presence of Streptococcus to adhere to the oral biofilm (plaque) and prefers lactate, the byproduct of metabolic process of *Streptococcus*, as its substrate of metabolism (3, 4). The top six genera present in this analysis: *Prevotella, Veillonella, Streptococcus, Actinomyces, Fusobacterium* and *Leptotrichia* represent abundant genera present in normal human oral microbiomes (5). Furthermore, the detection of low abundance microbes enables studies examining not only populations, but also active microbial evolution.

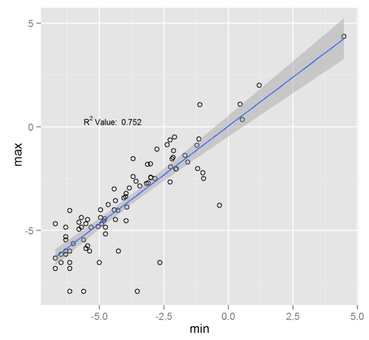

Figure 4. Correlation of two microbiome libraries produced using minimum PCR cycles – 16 and maximum PCR cycles – 28 from the same starting material

CONCLUSIONS

Microbial DNA enrichment reduced the number of off target reads, allowing detection of genera in low abundance in complex community studies.

The top ten represented genera in this study reflect proportions expected to be found in oral microbiome of a healthy individual.

Increased numbers of PCR cycles might be necessary for samples of low concentration or integrity. This does not negatively effect or influence sequencing results.

Precise and sensitive solution for large scale microbiome studies utilizing 384 unique 12 base sample indices.

Reliable taxonomic data from as low as 1 ng of bacterial DNA in as little as two hours hands on time.

REFERENCES

1. Lazarevic et al., Analysis of the salivary microbiome using culture independent techniques. Journal of Clinical Bioinformatics. 2012, 2:4.
2. Meyer et al., The Metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008, 9:386.
3. Kreth et al., Bacterial and Host Interactions of Oral *Streptococci*. DNA and Cell Biology. 2009, 28:8.
4. Distler, W., and Kroncke, A. The lactate metabolism of the oral bacterium *Veillonella* from human saliva. Arch Oral Biology. 1981, 26.
5. Dewhirst et al., The Human Oral Microbiome. Journal of Bacteriology. 2010, 192:19.
6. Marcel Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, (2011) EMBnet.journal, 17:10-12