High-Speed, Multiplexed 16S Microbial Sequencing on the MiSeq® System

Introduction

Next-generation sequencing is emerging as a powerful tool for profiling complex microbial communities. The MiSeq system provides the easiest sequencing workflow available, enabling researchers to go from sample to analyzed data in as little as eight hours. Using proven sequencing by synthesis (SBS) technology, MiSeq delivers > 1 gigabase (Gb) of data with quality comparable to that achieved with the HiSeq® 2000 platform. This application note describes sequencing 24 amplicon samples in a single MiSeq run using an indexing strategy and overlapping 2 × 150 bp reads.

Known differences between host-associated and free-living microbial communities and the biological conclusions drawn from sequence data should be reproducible across sequencing systems[1,2]. Post-sequencing microbial community analysis was performed using the QIIME (Quantitative Insights Into Microbial Ecology) pipeline[3], an open-source software package that integrates multiple standard community analysis tools. These results demonstrate that high-throughput microbial community sequencing previously performed on the Illumina Genome Analyzer™ can be successfully adapted for the MiSeq system, further decreasing costs and simplifying the workflow.

Methods and Results

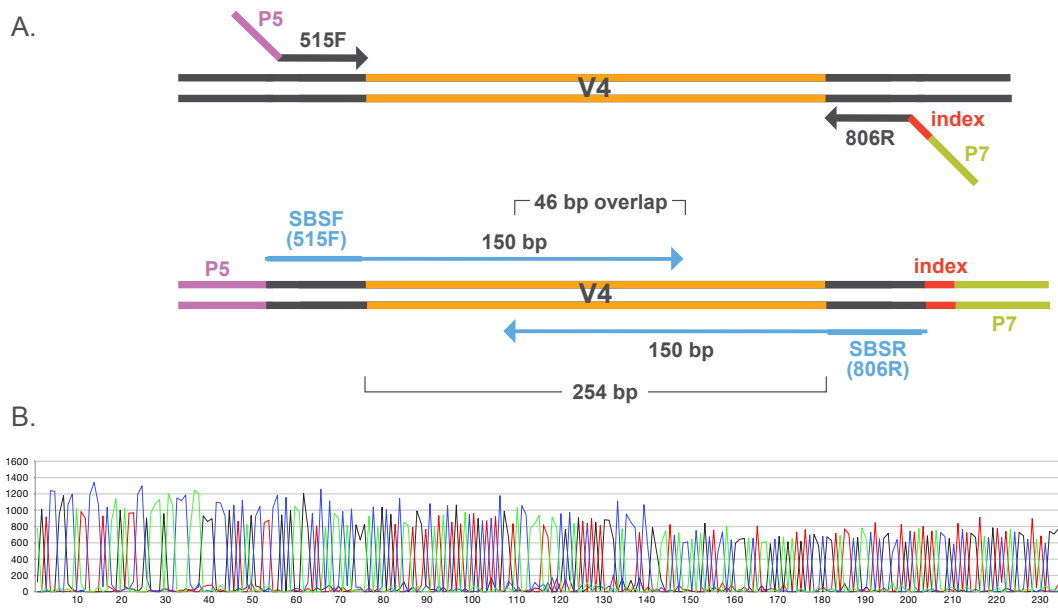
Sequencing the V4 region of 16S

Sequence variation in the 16S ribosomal RNA (rRNA) gene is widely used to characterize taxonomic diversity present in microbial communities. The 16S sequence is composed of nine hypervariable regions interspersed with conserved regions. The sequence of the 16S rRNA gene and its hypervariable regions has been determined for a large number of organisms, and is available from multiple databases such as Greengenes[4] and the Ribosomal Database Project[5]. For taxonomic classification, it is sufficient to sequence individual hypervariable regions instead of the entire gene length[6,7]. In most microbial species, the 16S fourth hypervariable (V4) region is approximately 254 bp, and only deviates from this length by a few basepairs.

V4 Amplification Strategy

To sequence the 16S V4 region, primers were designed against the surrounding conserved regions[2]. Because MiSeq enables paired 150-bp reads, the ends of each read were overlapped to generate extremely high-quality, full-length reads of the entire V4 region in a single overnight run (Figure 1). These primers were tailed with sequences to incorporate Illumina adapters with indexing barcodes. The V4 region from 24 samples was amplified using primers encoding

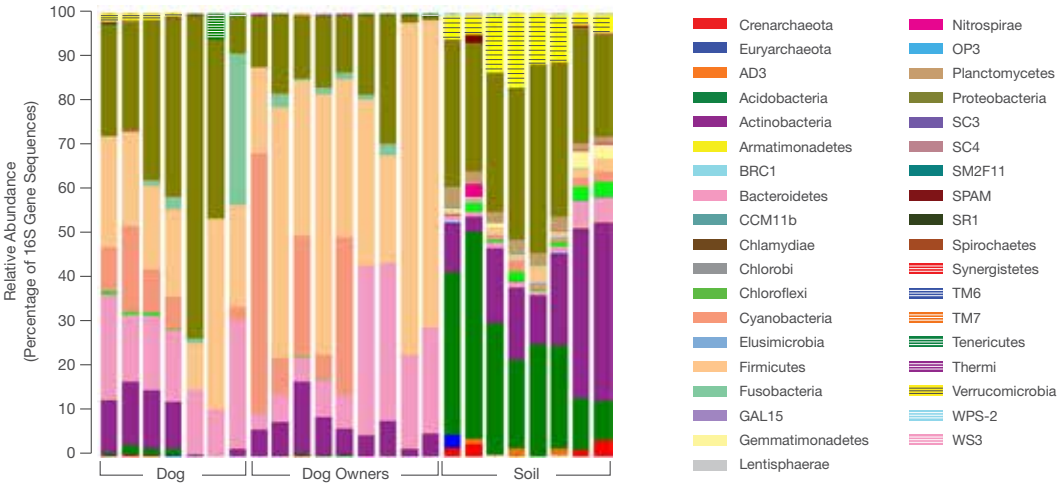
Figure 1: Amplification Strategy and Perfect Paired-End Read



A. V4 was amplified from each sample using primers 515F and 806R tailed with P5 and P7 sequences, respectively. Paired 150 bp sequencing gives a full-length 254 bp fragment of V4 with a 46 bp overlap. B. Raw intensities (matrix and phasing corrected) for an example perfect 254 bp paired-end read from the V4 library.

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTCGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGCCGGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA

Figure 2: QIIME Taxon Assignment at Phylum Level



To examine whether known differences in the 16S sequence between microbial communities could be reproduced using massively parallel sequencing, qseq files for read 1 were analyzed using QIIME. QIIME taxon assignment at phylum level for samples taken from various anatomic sites of a dog, the dog owners, and soil. Taxonomy assignments were made with QIIME using the Greengenes taxonomy.

24 different barcodes and combined into a single library for sequencing on the MiSeq system.

Sequencing on the MiSeq System

The sample containing 24 pooled barcoded samples was loaded onto the MiSeq reagent cartridge, and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with a 13-cycle index read was carried out without any further user intervention, taking 28 hours.

Data Analysis

Primary analysis (image analysis, basecalling) was performed on the MiSeq instrument. Quality filtered qseq files were analyzed offline using QIIME[8]. QIIME is designed to take users from their raw sequence data through to publication-quality graphics, including providing supporting analyses such as quality filtering of reads, demultiplexing, operational taxonomic unit (OTU) picking, taxonomy assignment, and alpha and beta diversity analyses. Figure 2 shows a phylum-level taxonomic summary based on read 1 (i.e., the 5' read) from the 24 samples from various environments, including dog, the dog's human owners, and soil samples. These samples were also concurrently run on the HiSeq 2000 system, and all data were highly reproducible across different flow cell lanes and across sequencing platforms[2] (data not shown).

Conclusions

With a simple multiplexing strategy and high data yield, the MiSeq system is ideally suited to microbial profiling, enabling rapid turnaround from sample to answer. The V4 region of the 16S ribosomal RNA from various microbial populations was sequenced, allowing phylum-level identification. Using open source tools such as QIIME, complex community analysis can be carried out, and publication-ready data available in a matter of days after sample acquisition. The MiSeq system has the capacity to accommodate a greater number of samples than those presented in this study, as well as multiple 16S variable regions, permitting deeper genomic scrutiny of larger metagenomic populations.

References

1. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA 108:4516–4522.
2. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, et al. (2011) Manuscript in preparation.
3. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman, FD, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7(5):335–336.
4. Desantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72(7): 5069–5072.
5. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucl Acids Res 37:D141–D145.
6. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. Nucl Acids Res 35:18.
7. Liu Z, DeSantis TZ, Andersen GL, Knight R. (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucl Acids Res 36:18.
8. http://qiime.org/tutorials/processing_illumina_data.html

Learn More

Go to www.illumina.com/miseq to learn more about the next revolution in personal sequencing.

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

© 2011-2012 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAIIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, SeqMonitor, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub. No. 770-2011-013 Current as of 06 February 2012

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA