illumina

High-Speed, Multiplexed 16S Amplicon Sequencing on the MiSeq[®] System

Get genus-level identification of complex microbial populations in 2 days with our simple workflow.

Introduction

Microbial populations thrive in a variety of environments, from deep forest soil to thermal vents on the ocean floor. Entire communities of microbes inhabit internal and external surfaces of humans and animals. Metagenomics is the study of metagenomes, or genetic material isolated from these microbial communities. Metagenomic surveys are commonly performed using the prokaryotic16S ribosomal RNA (rRNA) gene, which is approximately 1,500 bp and contains nine interspersed conserved and variable regions that facilitate sequencing and phylogenetic classification. Using next-generation sequencing to focus on targeted regions of the 16S rRNA gene, researchers can identify species in diverse microbial populations (Figure 1).

This application note describes a method and results of performing 16S amplicon sequencing on the MiSeq benchtop sequencing system. Building on Illumina's established amplicon sequencing protocol, users design primers that target a single region or multiple genomic regions, depending on the level of identity desired. This method can be used for any targeted amplicon sequencing, such as virus research, mutation detection, or other microbiologyrelated studies.

Methods

Targeted Amplification Strategy

Illumina's upcoming demonstrated protocol for 16S rRNA sequencing¹ targets the V3 and V4 regions (an amplicon of 459 bp, although length varies depending on the organisms) of the 16S gene (Figure 2). Primers were designed against the surrounding conserved regions. Because MiSeq enables paired 250-bp reads, the ends of each read are overlapped to generate extremely high-quality, full-length reads of the entire V3 and V4 region in a single 40-hour run. Using a limited cycle PCR, Illumina sequencing adapters and dual-index barcodes are added to the amplicon. Libraries are then normalized and pooled prior to sequencing. This approach can be used to target various regions of the 16S gene² or any other gene of interest. Up to 96 samples can be multiplexed into a single library for sequencing on the MiSeq System.³

Sequencing

The sample containing indexed amplicons 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 dual index reads was performed. Total run time for this 2 x 250 bp run was 40 hours.

Data Analysis

Researchers can perform primary analysis (image analysis through basecalling) directly on the MiSeq instrument. MiSeq Reporter (MSR) is Illumina's free software that performs downstream analysis after a run is complete. It can be used on the MiSeq System, in the BaseSpace®



cloud environment, or on a standalone computer. MSR reports data down to the genus level of a metagenomic population, showing the percentage of each genus present.

Results

The sequence of the 16S rRNA gene and its hypervariable regions has been determined for a large number of organisms. These sequences are available from multiple databases, such as Greengenes⁴ and the Ribosomal Database Project.⁵ For taxonomic classification, it is sufficient to sequence hypervariable regions instead of the entire gene length.^{6,7} The MSR Metagenomics 16S rRNA workflow uses the Greengenes database to perform taxonomic classification.

MSR supports a 16S workflow and delivers genus-level identification. The 16S workflow uses a Bayesian classification method⁸ to assign taxonomics and a confidence score to each read. For any given taxonomic level, reads falling below a certain confidence score will not be classified, leading to a certain number of "unclassified" reads. In Figure 3, the graphical output from MSR shows the percentage of each genus present in the sample. Figure 3A shows the genus-level classification of organisms in a human saliva sample. Figure 3B depicts the genus-level classification for a control metagenomics sample containing equal proportions of DNA from *Bacillus cereus*, *Klebsiella pneumonia*, *Listeria innocua*, *Listeria monocytogenes*, *Proteus*



mirabilis, Proteus vulgaris, Pseudomonas aeruginosa, Pseudomonas stutzeri, Staphylococcus aureus, and Staphylococcus epidermidis.

Conclusions

Illumina's demonstrated 16S rRNA protocol can be used to sequence the V3 and V4 regions of the 16S rRNA gene on the MiSeq System, enabling genus-level identification of complex microbial populations. Simple multiplexing and increased data output allow rapid turnaround from sample preparation to results in less than 2 days. Using Illumina's MiSeq Reporter, researchers can perform complex community analysis and obtain publication-ready data. The MiSeq system can accommodate a large number of samples as well as multiple 16S regions, permitting deeper genomic scrutiny of larger metagenomic populations. Future improvements to the MSR software will feature enhanced reporting and visualization, and will include species-level sensitivity. This protocol is not limited to sequencing 16S amplicons; the workflow described here and in the demonstrated protocol document can be applied to other amplicon sequencing projects.

Other third-party tools are available for analyzing metagenomics samples such as Quantitiative Insights into Microbial Ecology (QIIME).⁹⁻¹⁰



Output from the 16S rRNA Metagenomics workflow on MiSeq Reporter (MSR), showing all genera represented in the test (A) and control (B) microbial samples. DNA samples courtesy of Cleber Ouverney from San Jose State University.

References

1. Illumina-demonstrated protocol for 16S Amplicon Sequencing

2. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, et al. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41(1).

3. 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.

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. Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007) Naïve Bayesian classifier for rapid assignment of rRNA Sequences into the new bacterial taxonomy. Appl Environ Microbiol 73(16):5261-5267.

9. 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.

10. 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.

This application note is being provided for informational purposes only. In some cases reagents may need to be purchased from non-authorized third party suppliers. Illumina does not guarantee nor promise technical support for the performance of our products used with reagents purchased from a non-authorized third party supplier.

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

FOR RESEARCH USE ONLY

© 2013 Illumina, Inc. All rights reserved.

Illumina, Illumina/Dx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, 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. 1270-2013-001 Current as of 28 May 2013

illumina