16S metagenomics sequencing with the iSeq™ 100 System

Fast and efficient microbial sequencing on the most affordable Illumina sequencing system

Highlights

- Culture-free, NGS-based microbial analysis
  Identify and compare bacterial populations from diverse microbiomes
- Cost-efficient microbial metagenomics
  Study bacterial populations quickly and affordably
- Simple one-button data analysis
  Analyze sequencing data easily with the 16S Metagenomics BaseSpace™ App

Introduction

Metagenomic studies are commonly performed by analyzing the prokaryotic 16S ribosomal RNA (16S rRNA) gene, which is approximately 1500 bp long and contains nine variable regions interspersed between conserved regions. Variable regions of 16S rRNA are frequently used for phylogenetic classification of genus or species in diverse microbial populations.[1–4]

The choice and number of 16S rRNA regions to sequence are areas of debate, and the region of interest might vary depending on requirements such as experimental objectives, design, and sample type. This application note describes a comprehensive workflow that combines the Illumina demonstrated protocol for 16S metagenomics sequencing (Part # 15044223) with the iSeq 100 System (Figure 1) and secondary analysis using BaseSpace Sequence Hub.

Figure 1: The iSeq 100 System–The iSeq 100 System harnesses the power of NGS in the most affordable, compact benchtop sequencing system in the Illumina portfolio.

Simple, integrated workflow

16S metagenomics sequencing on the iSeq 100 System is part of an integrated next-generation sequencing (NGS) workflow that includes library preparation of the 16S V3 and V4 amplicon, proven high-quality Illumina sequencing, and push-button data analysis in BaseSpace Sequence Hub (Figure 2). The entire workflow proceeds from DNA to data in less than 30 hours.

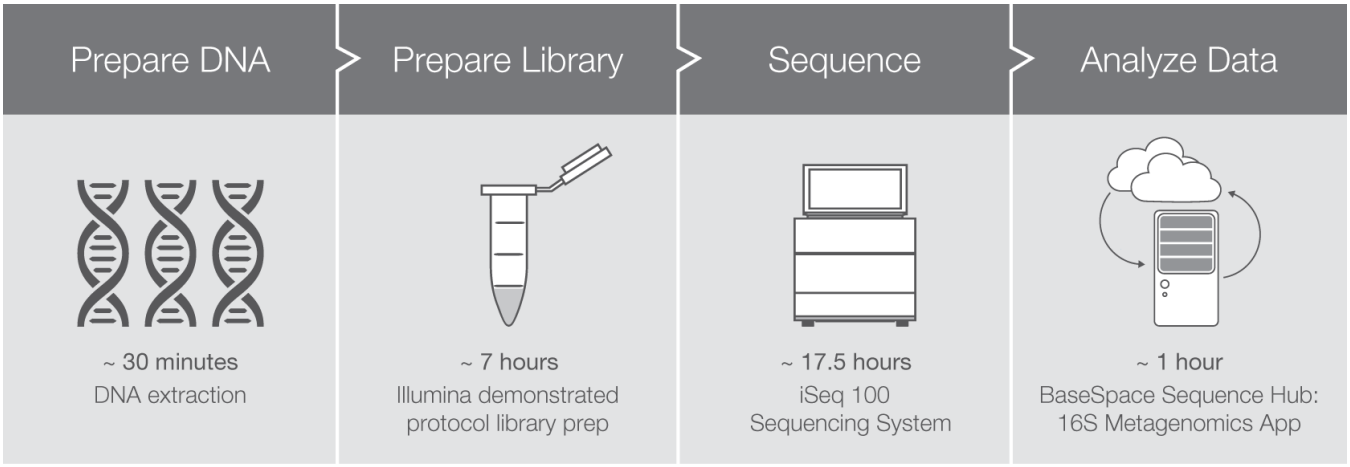

Figure 2: 16S metagenomics sequencing workflow–16S metagenomics sequencing on the iSeq 100 System is part of a streamlined, comprehensive NGS workflow that includes library preparation, sequencing, and data analysis.

Table 1: Primer sequences for 16S metagenomics sequencing

Name	Sequence[a]
16S amplicon PCR forward primer[b]	5′- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG - CCTACGGGNGGCWGCAG -3′
16S amplicon PCR reverse primer[b]	5′- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG - GACTACHVGGGTATCTAATCC -3′

a. International Union of Pure and Applied Chemistry (IUPAC) nucleotide nomenclature: N = any base; W = A or T; H = A or C or T; V = A or C or G
b. Primer sequence before the hypen is Illumina overhang adapter sequence. Primer sequence after the hyphen corresponds to locus-specific sequence.

Library preparation

The 16S metagenomics sequencing workflow begins with PCR amplification of the V3 and V4 regions of the 16S rRNA gene using a bacterial primer pair selected from the scientific literature (Table 1).[5] Illumina sequencing adapters and dual-index barcodes are then added to the generated amplicons using the Nextera™ XT DNA Index Kit. Libraries are normalized and pooled, and are ready for sequencing (Figure 3).

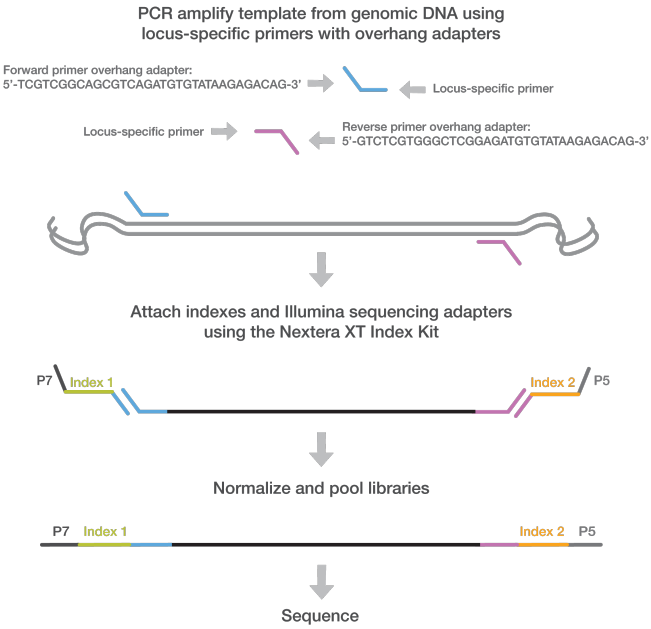

Figure 3: 16S V3 and V4 amplicon chemistry–Locus-specific primers with Illumina overhang adapters are used to amplify the V3 and V4 region of the 16S rRNA gene from genomic DNA. Sequencing adapters and dual-index barcodes are added, and libraries are normalized and pooled before sequencing.

Sequencing on the iSeq 100 System

After preparation, libraries are loaded into a prefilled reagent cartridge for sequencing on the iSeq 100 System. Starting a run on the iSeq 100 System is as easy as load and go with less than five minutes of setup. The iSeq 100 System integrates clonal amplification on a single instrument. The intuitive user interface provides guidance through every step of the run setup and run initiation processes, allowing researchers to perform various sequencing applications with minimal user training and set up time. The iSeq 100 System harnesses proven Illumina SBS chemistry, used to generate more than 90% of the world's sequencing data.[6] Illumina SBS chemistry is used in all Illumina sequencing systems, enabling researchers to compare data across systems and scale their studies to higher throughput systems.

Easy, flexible data analysis

Sequence data can be instantly transferred, analyzed, and stored securely in BaseSpace Sequence Hub, the Illumina genomics computing environment. BaseSpace Sequence Hub features a rich ecosystem of commercial and open-source apps for downstream data analysis. The 16S Metagenomics App performs taxonomic classification of 16S rRNA targeted amplicon reads using a version of the GreenGenes taxonomic database curated by Illumina (Table 2).

Table 2: 16S Metagenomics BaseSpace App

BaseSpace App	Description
16S Metagenomics App	Sample comparisons can be performed using the 16S Metagenomics App which enables analysis of 16S rRNA amplicon sequencing data and provides interactive visualization of taxonomic classification and relative abundance.

Experimental methods and results

To demonstrate the exceptional performance of the iSeq 100 System as part of a demonstrated protocol for 16S rRNA amplicon sequencing, data generated on the iSeq 100 System was compared against data generated on the MiSeq™ System for bacterial classification and relative abundance.

Methods

Samples and library preparation

Microbial genomic DNA samples were obtained from two sources for library prep, sequencing, and analysis. One source was the American Type Culture Collection (ATCC) 20 Strain Staggered Mix Genomic Material (ATCC MSA-1003). This mock microbial community comprises a staggered distribution of genomic DNA prepared from bacterial strains that were selected based on relevant attributes such as Gram stain, GC content values, and sporulation attributes. Real-world environmental samples were also obtained as part of a collaboration with academic researchers at the University of California, San Diego. Libraries were prepared following the 16S metagenomic sequencing library preparation workflow.[7] Prepared libraries were normalized and pooled before sequencing.

Table 3: Comparison of multiplexing capacity by sequencing system

Sequencing system	Multiplexing capacity[a]			Run data quality	
	PF paired reads[b]	15K reads per sample	100K reads per sample	% PhiX	% Reads ≥ Q30[c]
iSeq 100 System[d] 2 × 150 bp	4M	267	40	5	94.1
MiSeq System[e] 2 × 300 bp	25M	1667	250	10-25	74.9

a. Based on recommended 15K-100K reads per sample for analysis with 16S Metagenomics BaseSpace App.
b. Based on published instrument specifications.
c. Average of Read 1 and Read 2 data.
d. iSeq 100 System: v1 > 80% bases higher than Q30 at 2 × 150 bp.
e. MiSeq System: v3 > 70% bases higher than Q30 at 2 × 300 bp.

Sequencing and data analysis

Prepared and pooled libraries were run at varying read lengths on the iSeq 100 and MiSeq Systems. Sequencing results were analyzed using the 16S Metagenomics App in BaseSpace Sequence Hub.

Comparison of multiplexing capacity on the iSeq 100 System

The multiplexing capacity of the iSeq 100 and MiSeq Systems shows a high sample multiplexing ability across all instruments based on the need for 15K-100K reads per sample for the 16S Metagenomics App in BaseSpace Sequence Hub. The iSeq 100 System is able to take advantage of the 384 indexes available with Nextera XT DNA Index kits (Table 3).

Comparable Q30 scores with the iSeq 100 System

Sequencing low-diversity libraries, such as those used for 16S rRNA sequencing, is challenging due to unbalanced base composition, causing a large percentage of the clusters to show the same base during each cycle. The high signals caused by the imbalance result in low Q-scores even though the base calling accuracy is not necessarily poor. Therefore, a 5% PhiX spike-in enables error rate calculations that allow verification of base calling accuracy over the course of the run, for all PhiX clusters, which can be extrapolated to the samples. Comparing the Q30 scores for the iSeq 100 System to the MiSeq System shows robust performance across all systems and run types, with the iSeq 100 System having higher Q30 scores while using less PhiX input (Table 3).

Quality score (Q-score): A metric in NGS that predicts or estimates the probability of an error in base calling. A Quality score (Q-score) serves as a compact way to communicate very small error probabilities. A high Q-score implies that a base call is more reliable and less likely to be incorrect.

Q30: A Q-score predicting that one in 1000 base calls will be incorrect. Q30 is widely considered a benchmark for high-quality data. A successful run will produce between 75-95% bases with Q30 scores or higher depending on the sequencing system, read length, and sequencing library quality.

Characterization of microbial composition across sequencing systems

To demonstrate the exceptional performance of the iSeq 100 System as part of a demonstrated workflow for 16S metagenomics sequencing, mock community and real-world samples were interrogated on the iSeq 100 and MiSeq Systems.

Characterization of ATCC Microbiome Standards

In order to compare performance across systems, the 20 Strain Staggered Mix Genomic Material (ATCC MSA-1003) was sequenced across multiple systems and run types. Analysis of the sequencing data with the 16S Metagenomics App on the iSeq 100 System identified all members of the bacterial community and showed comparable performance to the MiSeq System with fewer reads, shorter reads, and lower PhiX spike-in (Figure 4). The use of 2 × 300 bp read length for the MiSeq System showed equivalent identification of bacteria using the 16S Metagenomics App, demonstrating the robustness of the analysis tools with lower read lengths ideal for the iSeq 100 System (Figure 4).

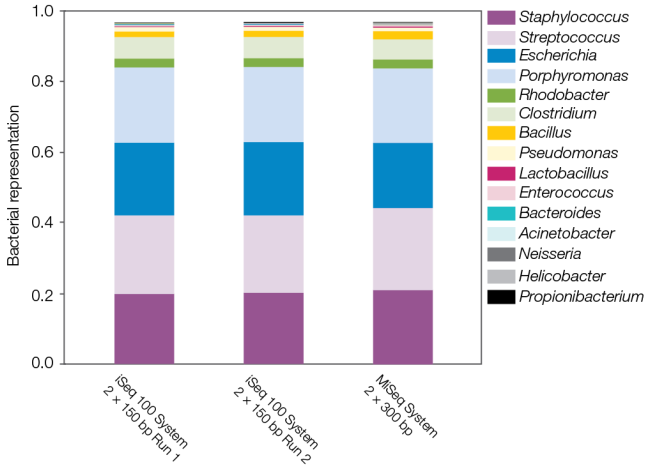

Figure 4: Comparative analysis of microbial composition of ATCC samples across systems—Analysis of microbial composition of ATCC samples with the iSeq 100 System results in excellent genera coverage as compared to the MiSeq System.

Characterization of real-world samples

Microbial composition of real world soil samples (Figure 5 A and B) and fecal samples (Figure 5 C and D) were compared using the 16S metagenomics sequencing workflow with the iSeq and MiSeq Systems. The community profiles of all samples tested were highly concordant between the iSeq 100 and MiSeq Systems (Figure 5). These results further reinforce the use of shorter read lengths on the iSeq 100 System for 16S metagenomics applications with real-world samples.

To further interrogate the real-world samples, the 10 highest represented genera were compared between Fecal Sample 1 and Soil Sample 1 to demonstrate the difference in bacterial identity as well as the difference in distribution for the highest genera found in these two sample types (Figure 6). No overlap is seen in the top represented genera between the two samples and the fecal samples show the bacterial community is more heavily dominated by a smaller number of bacterial genera.

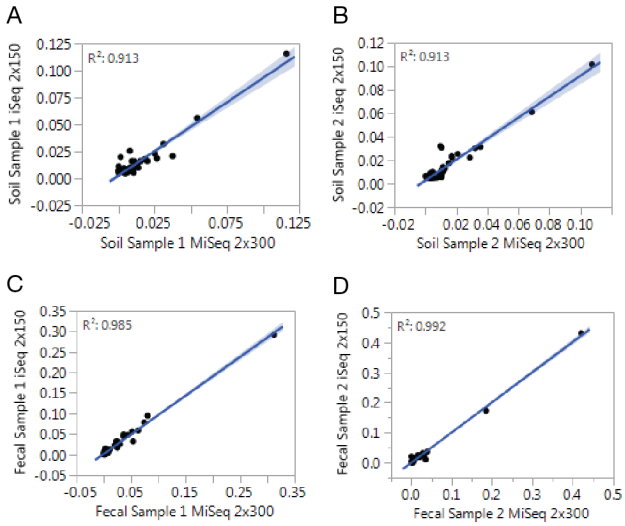

Figure 5: Comparative analysis of microbial composition of real world samples across sequencing systems–Analysis of fecal and soil samples for bacterial representation was highly concordant between the iSeq 100 and MiSeq Systems. Each axis is the fractional representation of each genera in each sample plotted against each other.

Summary

Using the 16S metagenomics workflow with the iSeq 100 System, microbiologists can achieve genus-level sensitivity for metagenomic surveys of bacterial populations. In this study, the Illumina workflow was used to study microbial populations in ATCC Microbiome Standards comprised of mock communities and in real-world samples. 16S metagenomic studies comprise one of many applications empowered by the iSeq 100 System. Illumina solutions support researchers during every step of the process, from DNA isolation through data analysis, enabling a range of applications for microbial genomics.

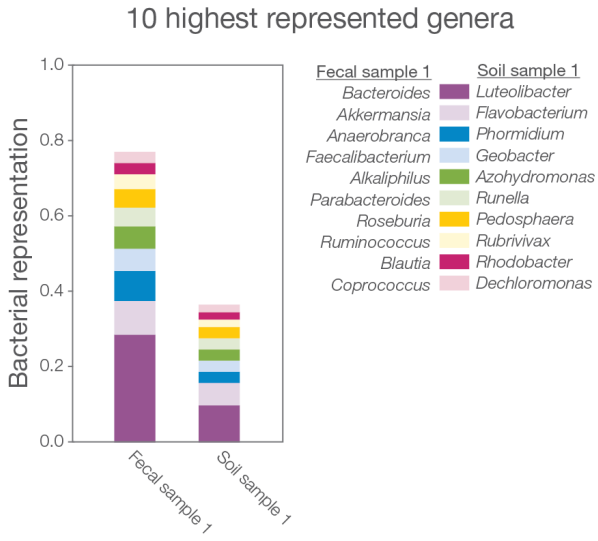

Figure 6: Analysis of bacterial populations in distinct microbiomes–Analysis of bacterial composition and distribution in fecal and soil samples showed no overlap in the top represented genera between the two samples. This confirms the samples are from two distinct microbiomes. **Note:** *Clostridium* was found in both fecal and soil samples tested but was not found to be one of the 10 highest represented genera in these samples.

Learn More

To learn more about the iSeq 100 System, visit www.illumina.com/systems/sequencing-platforms/iseq.html

To learn more about 16S metagenomics sequencing, visit www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/16s-rrna-sequencing.html

References

1. Woes CR, Fox GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *PNAS*. 1977;74(11):5088-5090.
2. Gray MW, Sankoff D, Cedergren JR. On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucleic Acids Res*. 1984;12 (14):5837-5852.
3. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *PNAS*. 1990;87(12):4576-4579.
4. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol*. 1991;173(2):697-703.
5. Klindworth A, Pruesse E, Schweer T, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. 2013;41(1):e1.
6. Data calculations on file. Illumina, Inc., 2017.
7. Illumina. 16S Metagenomci Sequencing Library Preparation User Guide. support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf. Accessed April 2018.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com