Comprehensive microbiology solutions

Influenza

16S rRNA

E. coli

Virome

MRSA

Zika

Culture-free

C. difficile

Ebola

Microbiome

and more...

Metagenomics

Illumina next-generation sequencing (NGS) offerings for microbiology

Area of Interest	Technology & Applications	Workflow: Library preparation	Sequencing options	Data analysis
• Antimicrobial resistance • Basic research • Epidemiology • Infectious disease	**Whole-genome sequencing** Generate accurate reference genomes for microbial identification and/or other comparative genomic studies starting from culture.	 Nextera™ DNA Flex, Nextera XT Kit, TruSeq™ DNA Kit	 iSeq™ System, MiniSeq™ System, MiSeq™ Series, and NextSeq™ Series	 Cloud: BaseSpace™ Sequence Hub On-Instrument: MiSeq Reporter
• Culture-free • Microbiome • Bacteria and archaea • Phylogenetic classifications	**16S rRNA sequencing** Identify and compare bacteria and archaea present within a given sample.	 16S library prep protocol with Nextera XT Index Kit	 iSeq System, MiniSeq System, MiSeq Series, NextSeq Series	 Cloud: BaseSpace Sequence Hub On-Instrument: MiSeq Reporter
• Community profiling • Culture-free • Environmental • Microbiome • Resistome • Virome	**Shotgun metagenomics** Identify, characterize, and detect low-abundance members of the microbial community.	 Nextera DNA Flex, Nextera XT Kit	 NextSeq Series, NovaSeq™ System	 Cloud: BaseSpace Sequence Hub
Virology • Pan-viral • Ebola • Zika • Influenza • and more...	**Shotgun metagenomics** **RNA sequencing** Gene expression, functional profiling, and viral sequencing.	 Nextera DNA Flex, Nextera XT Kit, TruSeq RNA Library Prep & Enrichment	 MiniSeq System, MiSeq Series, NextSeq Series, NovaSeq System	 Cloud: BaseSpace Sequence Hub

Data analysis & Bioinformatics solutions

BaseSpace™ Sequence Hub is a powerful cloud-based computing environment where NGS data can be easily stored, analyzed, and shared. Sequencing data from Illumina instruments are streamed in real-time over the internet to BaseSpace. The data is automatically converted to standard file formats (FASTQ) for analysis by applications (apps). Users launch apps in BaseSpace with just a few mouse clicks to analyze and visualize their data.

Whole-genome sequencing

Bacterial Analysis Pipeline: Identifies species, multilocus sequence type, plasmids, virulence, and antimicrobial resistance genes in bacteria.

SPAdes Genome Assembler: An open source tool for *de novo* sequencing. This app is designed to assemble small genomes.

SRST2: Reports sequence types from a MLST database and/or reference genes from a database of sequences for virulence genes, resistance genes, and plasmid replicons.

Velvet *de novo* Assembly: Use the Velvet assembler for *de novo* assembly of bacterial genomes.

16S rRNA sequencing

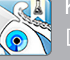
16S Metagenomics: Performs taxonomic classification of 16S rRNA targeted amplicon reads using an Illumina-curated version of the GreenGenes taxonomic database.

Kraken Metagenomics: Assigns taxonomic labels to short DNA sequences with high sensitivity and speed using exact alignments of k-mers and a novel classification algorithm.

QIIME: An open-source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data.

Shotgun metagenomics

GENIUS Metagenomics: Know Now: CosmosID's curated genome database provides rapid, accurate, and actionable bacterial identification at the species, subspecies, and/or strain level.


Kraken Metagenomics: Assigns taxonomic labels to short DNA sequences with high sensitivity and speed using exact alignments of k-mers and a novel classification algorithm.

MetaPhlAn (metagenomic phylogenetic analysis): A computational tool for profiling the composition of microbial communities from metagenomic shotgun sequencing data.

Virology

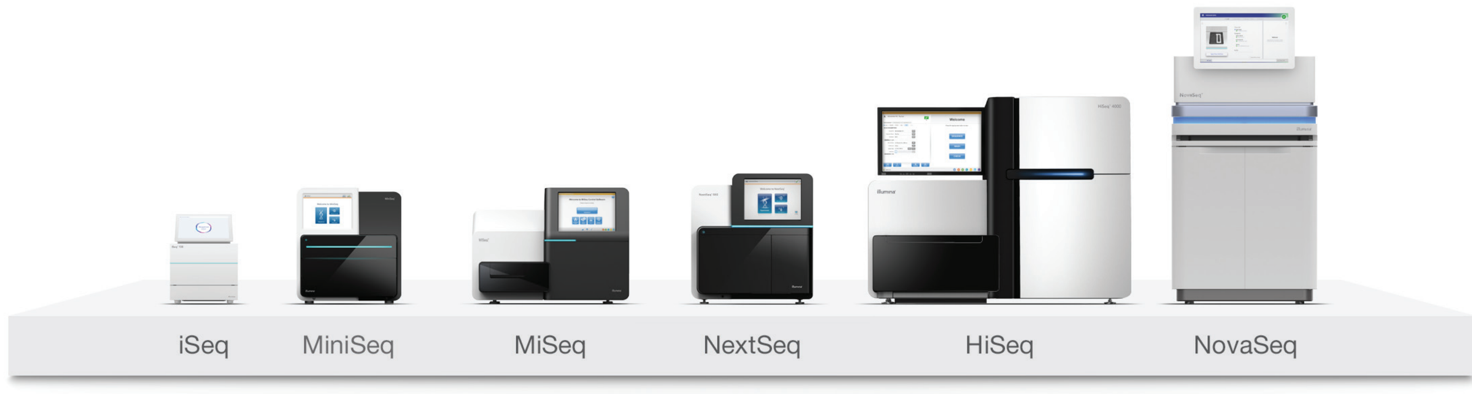
DeepChek® (HBV, HCV, HIV): Performs deep sequencing analyses and reports on subtyping, genotyping, and inferred levels of resistance.

Kraken Metagenomics: Assigns taxonomic labels to short DNA sequences with high sensitivity and speed using exact alignments of k-mers and a novel classification algorithm.

Getting started is easy

When you partner with Illumina, you become part of a community with more than 10,000* publications in microbiology and virology. Our 'starter bundle' packages provide training, library prep, sequencing instrument, and reagent kits that support a wide range of sample volumes. We also offer a program that allows you to trade in your sequencer for an Illumina system.

Explore the exciting discoveries our NGS solutions are enabling. Visit **www.illumina.com/MicroStories** and **www.illumina.com/MicroWebinars**.



*Illumina maintains an up-to-date database of all published scientific articles that use Illumina technology.

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

For Research Use Only. Not for use in diagnostic procedures.

© 2018 Illumina, Inc. All rights reserved. Pub. No. 1270-2016-003 QB 6390

Microbial Whole-Genome Sequencing with the iSeq™ 100 Sequencing System

Fast and efficient sequencing that provides uniform coverage and genome assembly for microbial species.

Highlights

- **Streamlined Workflow**
  Access a comprehensive workflow from DNA to data
- **Optimized Library Prep**
  Obtain robust, consistent results over a wide range of DNA input, even at low DNA input amounts (1 ng)
- **Comprehensive Coverage**
  Produce sequencing data with uniform coverage for viruses, bacteria, and other microbes

Introduction

Next-generation sequencing (NGS) has been established as an important tool in microbiology research for analysis of small genomes (≤ 5 Mb), including bacteria, viruses, and other microbes. Microbial NGS, including whole-genome sequencing (WGS) and targeted resequencing, enables mapping and *de novo* assembly of novel organisms, completing genomes of known organisms, and comparing genomes across samples.

The development of Nextera™ chemistry shortened and simplified library preparation by consolidating DNA fragmentation and adapter tagging steps into a single reaction (termed tagmentation) and eliminating the need for library quantitation before pooling and sequencing.[1] The Nextera DNA Flex Library Preparation Kit represents the next step in the evolution of Illumina library prep. In addition to speed and efficiency gains in the workflow, Nextera DNA Flex offers exceptional flexibility for sample input type and amount and robust, consistent preparation of sequencing-ready libraries.

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.

The latest innovation in NGS is poised to advance microbiology genomics research. The compact iSeq 100 System (Figure 1) combines complementary metal-oxide–semiconductor (CMOS) technology with proven Illumina sequencing by synthesis (SBS) chemistry to deliver high-accuracy data with fast time to results. The iSeq 100 System is part of a streamlined NGS workflow for targeted and whole-genome microbial sequencing (Figure 2).

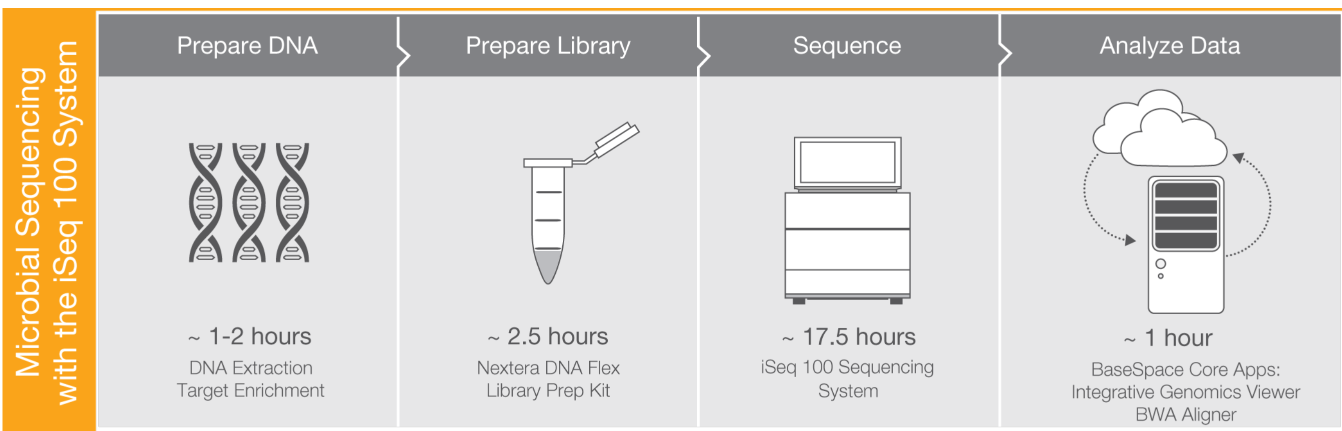


Figure 2: Microbial Sequencing Workflow—Microbial sequencing on the iSeq 100 System is part of a streamlined, comprehensive NGS workflow that includes Nextera DNA Flex library preparation, sequencing, and data analysis.

For Research Use Only. Not for use in diagnostic procedures.

Simple, Integrated Workflow

Microbial sequencing on the iSeq 100 System is part of an integrated NGS workflow that includes library preparation with the Nextera™ DNA Flex Library Preparation Kit, proven Illumina sequencing, and push-button data analysis in BaseSpace™ Sequence Hub (Figure 2). The entire workflow proceeds from DNA to data in less than 24 hours.

Optimized Library Prep

A major advance in Illumina library prep chemistry and key feature of the Nextera DNA Flex Library Preparation Kit is On-Bead Tagmentation, which uses bead-linked transposomes (BLTs) to mediate simultaneous DNA fragmentation and the tagging of Illumina sequencing primers (Figure 3).

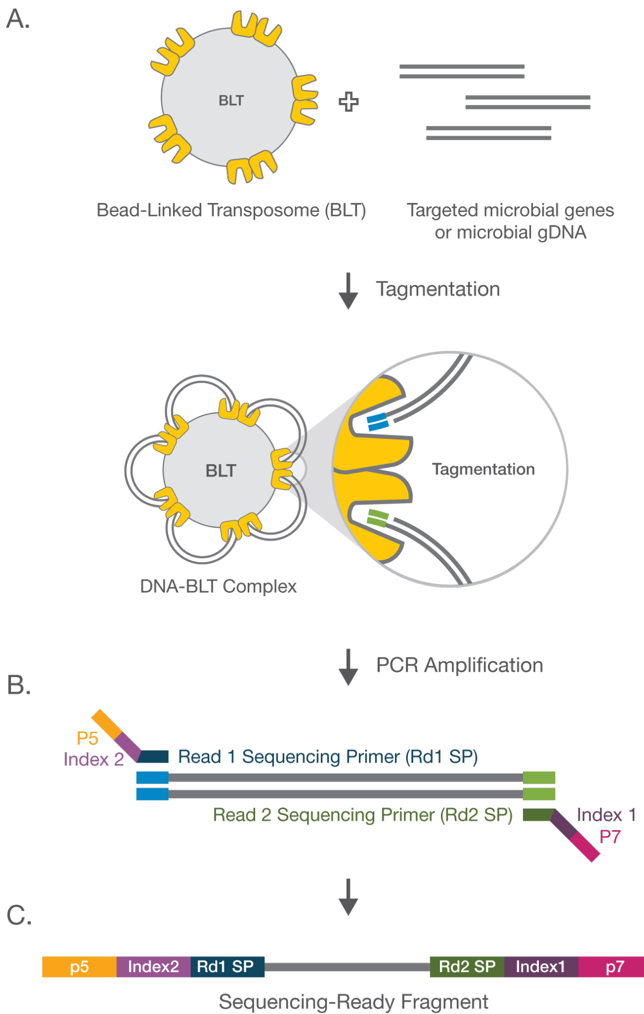


Figure 3: Nextera On-Bead Tagmentation Chemistry—(A) BLTs mediate tagmentation. (B) Reduced-cycle PCR amplifies sequencing ready DNA fragments and adds indexes and adapters. (C) Sequencing-ready fragments are washed and pooled.

To learn more about Nextera DNA Flex and On-Bead Tagmentation, read the Nextera DNA Flex Library Preparation Kit Data Sheet at www.illumina.com/nextera-dna-flex

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, sequencing, and data analysis into 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 minimal set up time.

The iSeq 100 System harnesses proven Illumina SBS chemistry, used to generate more than 90% of the world's sequencing data.[2] 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

The iSeq 100 System offers several data analysis options, including onboard and cloud-based data analysis. The Local Run Manager software, an onboard analysis software, features modular architecture to support current and future assays. Local Run Manager software supports planning sequencing runs, tracking libraries and runs with audit trails, and integration with onboard data analysis modules.

Alternatively, 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, including the Integrative Genomics Viewer and BWA Aligner apps (Table 1).

Table 1: BaseSpace Apps for MIcrobial Sequencing Data Analysis

BaseSpace App		Description
igv	Integrative Genomics Viewer	The Integrative Genomics Viewer (IGV) app displays alignments and variants from multiple samples for performing complex variant analysis.
	BWA Aligner	The BWA Aligner app aligns samples (FASTQ files) to a reference genome using the Burrows-Wheeler Aligner maximal exact match (BWA-MEM) algorithm.

Comprehensive Coverage

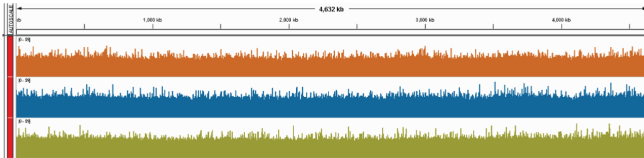
To demonstrate the comparable performance of the iSeq 100 System to other sequencing systems in the Illumina portfolio in the genome assembly of microbial organisms, input genomic DNA from three different bacterial species with varying GC content (Table 2) were prepared with the Nextera DNA Flex Library Preparation Kit. Libraries were sequenced using paired-end 2 × 151 bp reads on the iSeq 100 System, MiniSeq™ System, and MiSeq™ System.

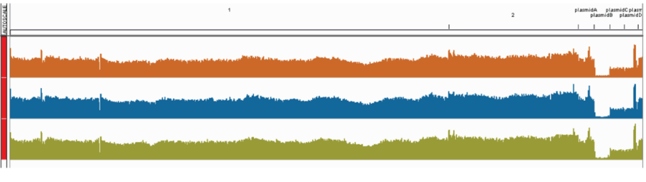
The iSeq 100 System delivers similar uniformity of coverage across different bacterial species, as compared to the MiniSeq and MiSeq Systems (Figure 4). These results support the exceptional performance of the iSeq 100 System for targeted and whole-genome microbial sequencing.

Table 2: GC Content of Sequenced Microbial Genomes

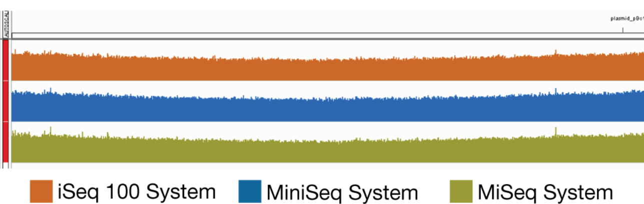
	B. cereus	*E. coli*	*R. sphaeroides*
Genome Size	~ 5.4 Mb	~ 4.6 Mb	~ 4.1 Mb
% GC Content	~ 35%	~ 51%	~ 69%

Escherichia coli





Bacillus cereus



iSeq 100 System MiniSeq System MiSeq System

Figure 4: Consistent Uniformity of Coverage —The iSeq 100 System delivers similar uniformity of coverage across different three bacterial species, as compared to the MiniSeq and MiSeq Systems.

Summary

The iSeq 100 System is part of a fully supported solution for targeted and whole-genome microbial sequencing that includes simplified library preparation with the Nextera DNA Flex Library Preparation Kit, sequencing, and user-friendly data analysis. The iSeq 100 System delivers the same data quality as larger benchtop sequencers in a smaller footprint with faster run times, making it an ideal, cost-effective solution for small-scale microbiology NGS applications.

Ordering Information

Library Prep	Catalog No.
Nextera DNA Flex Library Prep Kit (24 samples)	20018704
Nextera DNA Flex Library Prep Kit (96 samples)	20018705
Nextera DNA CD Indexes (24 indexes, 24 samples)	20018707
Nextera DNA CD Indexes (96 indexes, 96 samples)	20018708
Sequencing System	**Catalog No.**
iSeq 100 System	20021532
Sequencing Reagents	**Catalog No.**
iSeq 100 i1 Reagents (300 cycles single kit)	20021533
iSeq 100 i1 Reagents (300 cycles quad kit)	20021534

Learn More

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

To learn more about microbial whole-genome sequencing, visit www.illumina.com/microbiology.html

References

1. Illumina. Nextera XT DNA Library Preparation Kit Data Sheet. www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_nextera_xt_dna_sample_prep.pdf. Accessed September 2017.
2. Data calculations on file. Illumina, Inc., 2017.

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

© 2017 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see www.illumina.com/company/legal.html. Pub. No. 770-2017-034-A QB 5475

illumina®

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.

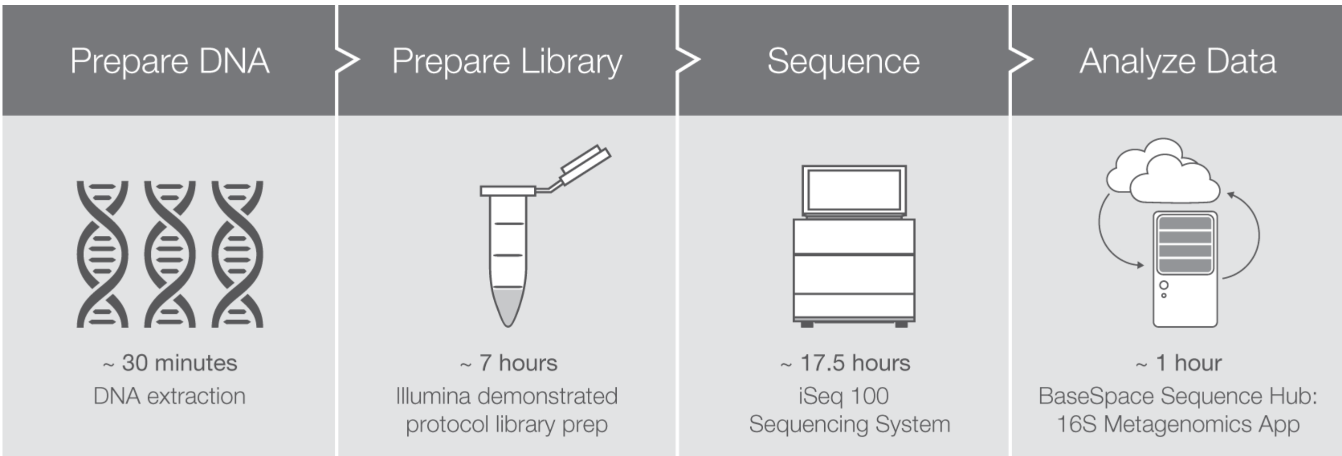


Figure 2: 16S metagenomics sequencing workflow—16S metagenomics sequencing on the iSeq 100 System is part of a streamlined, comprehensive NGS workflow that includes library preparation, sequencing, and data analysis.

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).

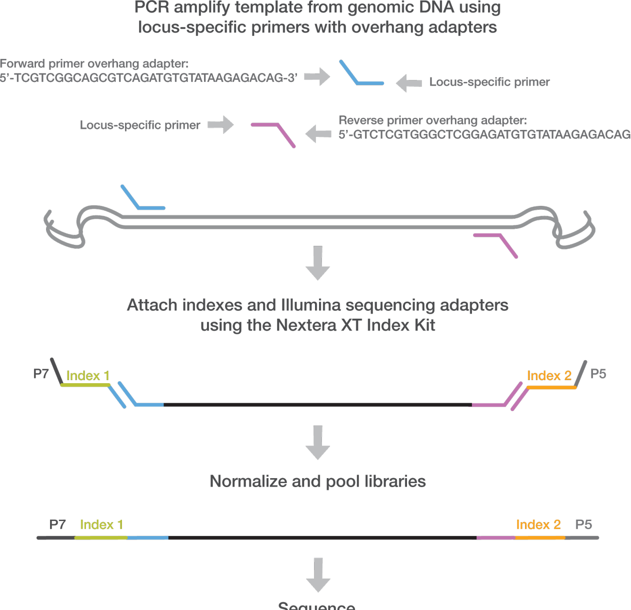


Figure 3: 16S V3 and V4 amplicon chemistry—Locus-specific primers with Illumina overhang adapters are used to amplify the V3 and V4 region of the 16S rRNA gene from genomic DNA. Sequencing adapters and dual-index barcodes are added, and libraries are normalized and pooled before sequencing.

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).

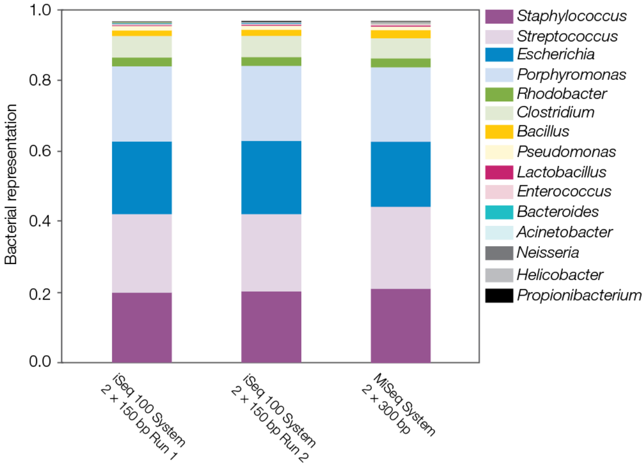


Figure 4: Comparative analysis of microbial composition of ATCC samples across systems—Analysis of microbial composition of ATCC samples with the iSeq 100 System results in excellent genera coverage as compared to the MiSeq System.

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.

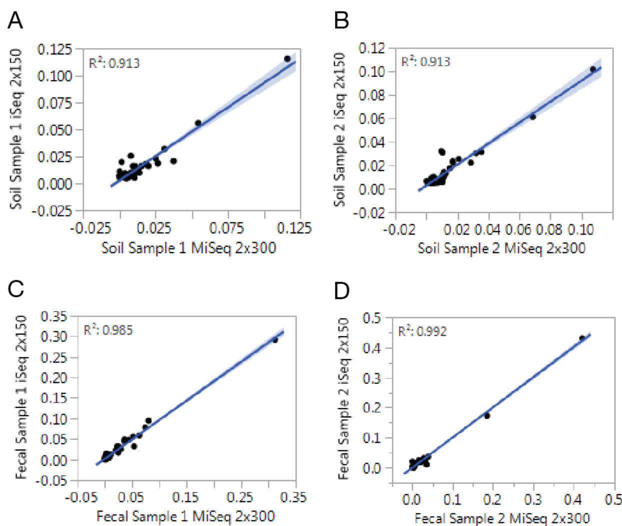


Figure 5: Comparative analysis of microbial composition of real world samples across sequencing systems—Analysis of fecal and soil samples for bacterial representation was highly concordant between the iSeq 100 and MiSeq Systems. Each axis is the fractional representation of each genera in each sample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

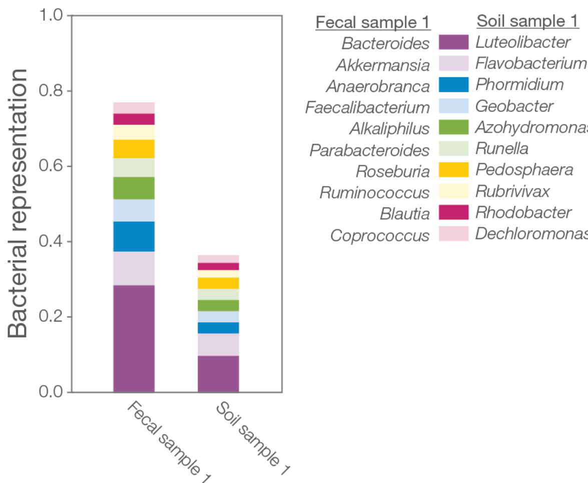


Figure 6: Analysis of bacterial populations in distinct microbiomes—Analysis of bacterial composition and distribution in fecal and soil samples showed no overlap in the top represented genera between the two samples. This confirms the samples are from two distinct microbiomes. **Note:** *Clostridium* 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

© 2018 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see www.illumina.com/company/legal.html. Pub. No. 770-2018-009-B QB# 6012

16S Metagenomics Studies with the MiSeq® System

From DNA isolation to analysis, this simple workflow enables species-level identification of complex microbial populations in 2 days.

Introduction

Metagenomic surveys of microbial populations are often performed using the prokaryotic 16S ribosomal RNA (rRNA) gene, which contains conserved and variable regions that facilitate sequencing and phylogenetic classification. Following the complete Illumina workflow (Figure 1), 16S metagenomics studies with the MiSeq System can achieve species-level identification of microbial populations efficiently. The workflow includes DNA isolation, library preparation, sequencing, and push-button analysis, delivering an end-to-end solution for 16S metagenomics. By combining the demonstrated Illumina library preparation protocol, the MiSeq System, and simple analysis software, researchers can analyze complex microbial samples quickly and easily.

This application note provides an overview of the Illumina 16S metagenomics workflow and the results of a study that examined microbial populations in a water reservoir. This study was conducted in collaboration with the Center for Earth and Environmental Science at Indiana University–Purdue University Indianapolis (IUPUI) and Citizens Energy Group. The results reveal patterns in the reservoir's microbial community that can potentially be used to assess environmental influences on water quality.

Methods

The study used the 16S Metagenomic Sequencing Library Preparation Guide[1] to prepare sequencing libraries targeting the variable V3 and V4 regions of the 16S rRNA gene. Paired-end sequencing was performed on the MiSeq System and data were analyzed using the 16S Metagenomics App in the BaseSpace® analysis environment.

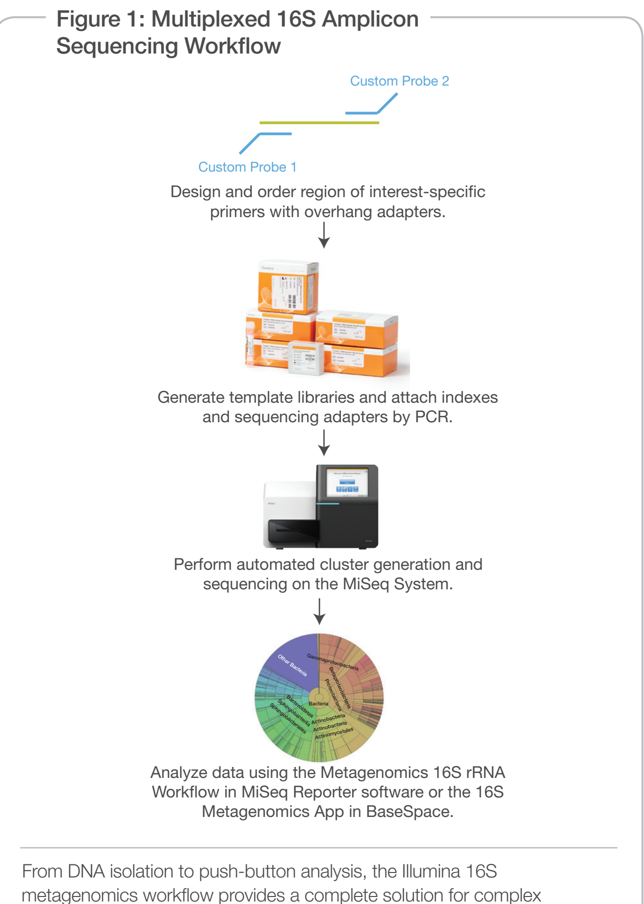
Sample Collection

Water samples were collected from the Eagle Creek reservoir in Indianapolis, Indiana at regular intervals throughout one year. Discrete samples were collected from the surface and at depths of 3 meters, 6 meters, and near the bottom floor of the reservoir.

DNA Isolation

Epicentre® DNA isolation kits deliver high-quality, inhibitor-free DNA from mixed samples of gram-positive and gram-negative bacteria derived from many environmental sources, including water, soil, fecal matter, and compost (Table 1). In this study, DNA was isolated from 27 water samples using the Meta-G-Nome™ DNA Isolation Kit[2]. Approximately 700 ng of DNA were extracted from each sample.

Figure 1: Multiplexed 16S Amplicon Sequencing Workflow



From DNA isolation to push-button analysis, the Illumina 16S metagenomics workflow provides a complete solution for complex community analyses. Sequencing with the MiSeq System delivers highly accurate data, and analysis includes classification using BaseSpace or MiSeq Reporter software.

Table 1: DNA Isolation Kits

Sample Type	Isolation Kit
Water	Metagenomic DNA Isolation Kit for Water[3]
Soil	SoilMaster™ DNA Extraction Kit[4]
Fecal matter	ExtractMaster™ Fecal DNA Extraction Kit[5]
Difficult-to-culture species present in environmental water, soil, or compost	Meta-G-Nome DNA Isolation Kit

Epicentre DNA isolation kits are optimized to isolate bacterial DNA obtained from various environments.

Table 2: MiSeq System Configurations

Flow Cell	No. of Reads	Read Length	Output	No. of 16S Samples Per Run
600-cycle V3 standard flow cell	25 M	2 × 300 bp	15 Gb	Hundreds of 16S samples
500-cycle V2 standard flow cell	15 M	2 × 250 bp	8 Gb	
300-cycle V2 micro flow cell	4 M	2 × 150 bp	1.2 Gb	Tens of 16S samples
500-cycle V2 nano flow cell	1 M	2 × 250 bp	0.5 Gb	

Library Preparation

The Illumina 16S Metagenomic Sequencing Library Preparation Guide is an easy-to-follow protocol for preparing DNA libraries. It is optimized to target the V3 and V4 regions of the 16S rRNA gene, although it can be adapted to target other variable regions. The 16S Metagenomic Sequencing Library Preparation Guide leads users through each step of library preparation, from genomic DNA to sequencing-ready libraries. All necessary reagents are listed, including the required primer sequences that target the V3 and V4 regions of the 16S rRNA gene. These primers can also be modified to target different regions of the 16S gene, or altered for custom applications. The 27 samples from the reservoir were prepared using the 16S library preparation protocol and the Nextera® XT DNA Index Kit[6] for cost-effective sample multiplexing.

Sequencing

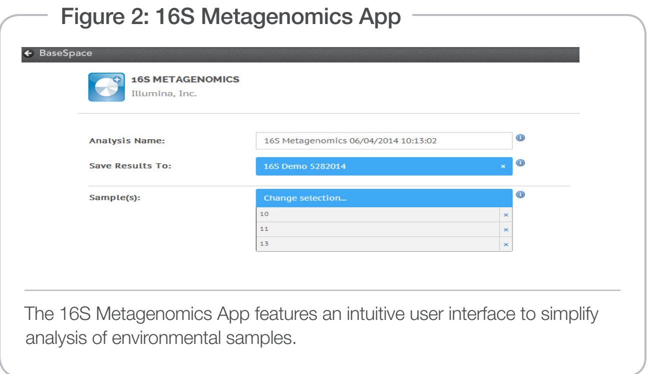
The MiSeq System can deliver 2 × 300 bp reads and up to 50 million paired-end reads, generating up to 15 Gb of data. The flexible system enables microbiologists to scale studies from one to hundreds of samples. Micro and nano flow cell options and accompanying reagents are available to support lower-throughput experiments by optimizing sample volume and coverage needs (Table 2).

Samples from the reservoir were loaded onto a MiSeq reagent cartridge and then onto the instrument. Automated cluster generation and a 2 × 300 bp paired-end sequencing run were performed. The resulting sequence reads were equally distributed across the samples, demonstrating uniform coverage.

Data Analysis

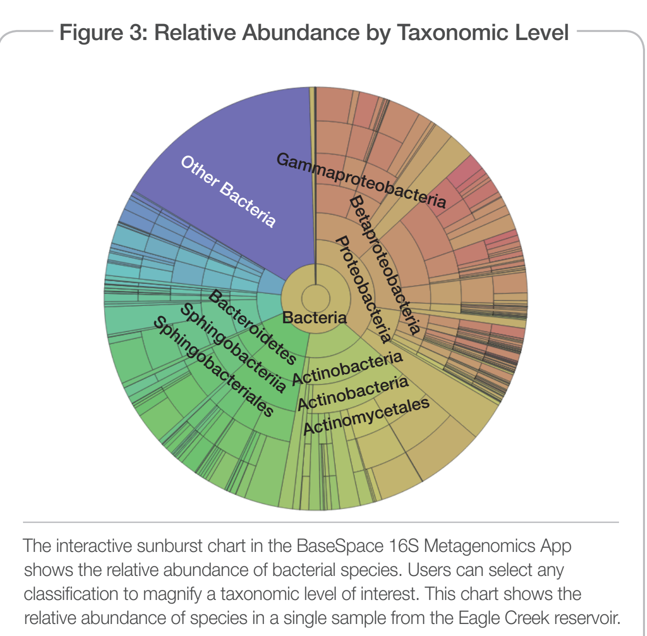
Illumina has removed much of the complexity from sequencing data analysis. Following the Illumina workflow, researchers can analyze sequencing data generated on the MiSeq System either on the instrument or in BaseSpace. MiSeq Reporter software is able to analyze data on the sequencer or on a standalone computer. Alternatively, data can be transferred, analyzed, stored, and shared with collaborators in BaseSpace. BaseSpace can deliver analyzed sequences in as little as 12 hours following the 16S workflow, and BaseSpace applications (apps) provide access to a growing collection of analysis tools.

Figure 2: 16S Metagenomics App



The 16S Metagenomics App features an intuitive user interface to simplify analysis of environmental samples.

Figure 3: Relative Abundance by Taxonomic Level



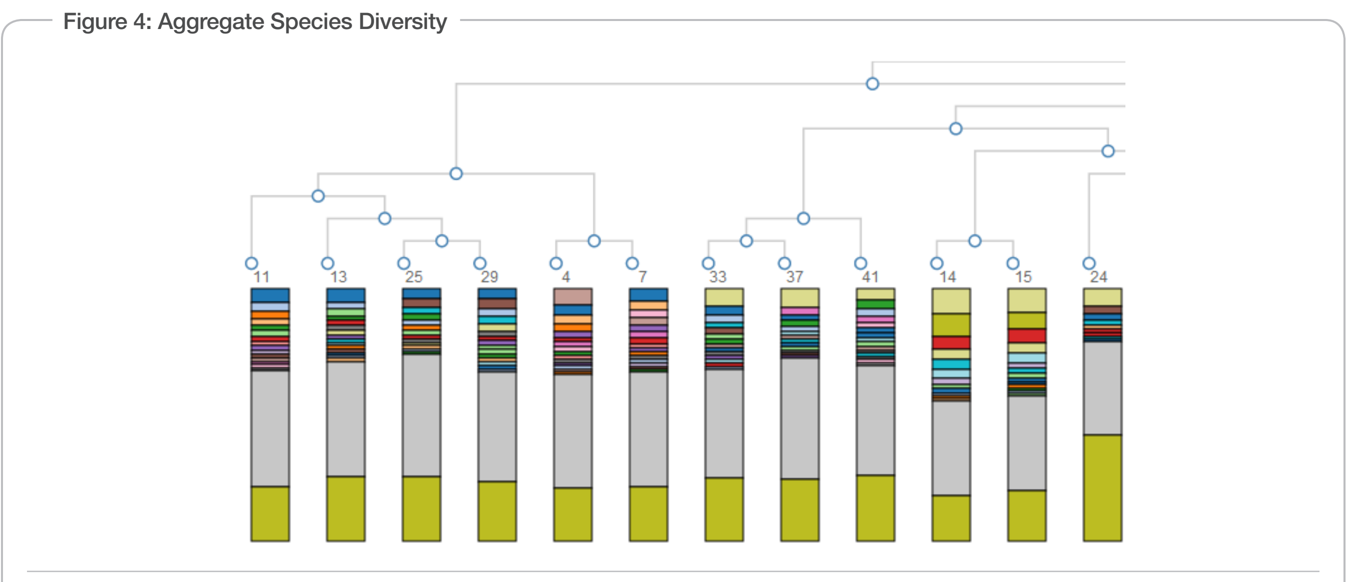
The interactive sunburst chart in the BaseSpace 16S Metagenomics App shows the relative abundance of bacterial species. Users can select any classification to magnify a taxonomic level of interest. This chart shows the relative abundance of species in a single sample from the Eagle Creek reservoir.

The reservoir samples were analyzed using the BaseSpace 16S Metagenomics App (Figure 2). The app delivers all phylogenetic data—including coverage statistics and detected species—in intuitive, easy-to-analyze reports. Sequencing reads are classified against the Greengenes[7] database, achieving up to species-level sensitivity.

Results

The 16S Metagenomics App delivers highly interactive visualizations for exploring taxonomic classifications. The sunburst classification chart provides a detailed view of the relative abundance of bacterial species within each taxonomic level. Researchers can select a category to magnify a particular level of interest and explore the diversity of any sample (Figure 3).

Figure 4: Aggregate Species Diversity



The 16S Metagenomics App delivers a dendrogram that shows a hierarchical clustering of samples based on genus-level classifications. The bar charts show the relative abundance of the classifications present in each sample. Each bar color corresponds to a unique genus.

The 16S Metagenomics App also provides an aggregate summary report so that researchers can compare the similarities and differences among samples within a project. The hierarchical dendrogram shows a clustering of samples based on genus-level classifications and the relative abundance of each (Figure 4). Detailed results and analysis from the Eagle Creek reservoir are available to view in BaseSpace[8]. Analysis of relative abundance in the reservoir revealed an increase in *Rhodococcus* species in July. Various factors may have caused the surge in *Rhodococcus* abundance, such as algaecide treatment or the use of fertilizer near the reservoir. Further analysis is required to assess the influences that contributed to changes in the community.

Conclusions

Using the 16S metagenomics workflow with the MiSeq System, microbiologists can achieve up to species-level sensitivity for metagenomic surveys of bacterial populations. In this study, the Illumina workflow was used to study microbial populations in a reservoir, uncovering shifts in community composition. This research enables biologists to investigate new methods for understanding environmental influences on water sources and water quality. 16S metagenomic studies comprise one of many applications empowered by the MiSeq desktop sequencer. Illumina solutions support researchers during every step of the process, from DNA isolation through data analysis, enabling a range of applications for microbial genomic discovery.

Learn More

For more information about the use of Illumina technology in microbial genomics, visit www.illumina.com/microbiology. To learn more about the MiSeq desktop sequencer, go to www.illumina.com/miseq.

References

1. Illumina (2013) 16S Metagenomic Sequencing Library Preparation Guide. (support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.ilmn)
2. Meta-G-Nome DNA Isolation Kit (www.epibio.com/applications/nucleic-acid-purification-extraction-kits/dna-purification-genomic/meta-g-nome-dna-isolation-kit) Accessed 07 July 2014.
3. Metagenomic DNA Isolation Kit for Water (www.epibio.com/applications/nucleic-acid-purification-extraction-kits/dna-purification-genomic/metagenomic-dna-isolation-kit-for-water) Accessed 07 July 2014.
4. SoilMaster- DNA Extraction Kit (www.epibio.com/applications/nucleic-acid-purification-extraction-kits/dna-extraction/soilmaster-dna-extraction-kit) Accessed 07 July 2014.
5. ExtractMaster- Fecal DNA Extraction Kit (www.epibio.com/applications/nucleic-acid-purification-extraction-kits/dna-extraction/extractmaster-fecal-dna-extraction-kit) Accessed 07 July 2014.
6. Nextera XT DNA Library Prep Kit | Illumina (www.illumina.com/products/nextera_xt_dna_sample_prep_kit.ilmn) Accessed 07 July 2014.
7. 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: 5069–72.
8. Project - BaseSpace (basespace.illumina.com/projects/9954952/MiSeq-v3-Nextera-XT-Metagenomics-Water-Samples) Accessed 07 July 2014.

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

FOR RESEARCH USE ONLY

© 2013–2014 Illumina, Inc. All rights reserved.
Illumina, BaseSpace, Epicentre, ExtractMaster, Meta-G-Nome, MiSeq, Nextera, SoilMaster, the pumpkin orange color, and the streaming bases design are trademarks or registered trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.
Pub. No. 1270-2013-001 Current as of 17 October 2017

Exploring the Microbial Communities Within and Around Us

Next-generation sequencing with the MiSeq® System enables researchers to study the microbiota of humans, model organisms, and clinical environments.

Introduction

We are never really alone. Each of us is a host to flourishing populations of microorganisms arranged in communities referred to collectively as microbiota. Researchers are finding that regardless of whether they're located in the gut, skin, or airways, these communities possess great diversity that can change as we age, in response to certain diseases, changes in diet, or the ingestion of therapeutic drugs. While some pathogenic microorganisms can lead to disease or even death, many are essential to human health and well-being.

New microbial profiling approaches, such as 16S ribosomal RNA (rRNA) sequencing on the MiSeq System, have led to a greater understanding of our microbial communities and their interactions with us. Christopher Taylor, PhD, is part of the Louisiana State University Health Science Center (LSUHSC) Microbial Genomics Resource Group, an organization that supports microbial genomics with scientific expertise and research services. As an Associate Professor in the Department of Microbiology, Immunology, and Parasitology, Dr. Taylor uses rRNA and DNA sequencing approaches to investigate microbes of importance to human health.

iCommunity spoke with Dr. Taylor about his microbiome projects and how the MiSeq System has enabled his studies.

Q: How did you become involved in metagenomic studies?
Christopher Taylor (CT): I have a computer science and mathematics background and got involved with computational biology when I was in graduate school, where I was part of the US National Human Genome Research Institute's Encyclopedia of DNA Elements (ENCODE) project[1-3]. I began my career as a faculty member at the University of New Orleans by focusing on applying high-throughput DNA and RNA sequencing in biological studies. One of my early projects was in collaboration with Dr. Erik Flemington, a virologist at Tulane University. We became interested in the RNA sequence reads from human cancer cell lines that did not map back to the human genome. In many labs at the time, the typical workflow was to map as many sequencing reads as possible back to the host genome, and then discard the remaining 15–20% of the reads. We wanted to look more closely at the nonmapping reads to see if we could find any viral, bacterial, or other recognizable sequences.[4-6] This is still an active collaboration, and our most recent paper shows that there is a lot of microbial contamination in existing RNA sequencing data sets.[7]

Now that I'm at the LSUHSC School of Medicine, there's more of a health care focus to my work, and I've become immersed in research on microbial communities. My primary focus over the last 4 years has been using 16S rRNA sequencing to study the different microbial communities that populate model organisms, humans, and the environments in which they live.

Q: What microbiomes are you studying?
CT: We have various ongoing studies looking at gut, vaginal, airway, and environmental microbiota. In a recent collaborative research study with Drs. Michael Ferris and Duna Penn at Children's Hospital of New Orleans, we used sequencing to look at the gut microbiota of infants in the neonatal intensive care unit, particularly premature infants suffering from necrotizing enterocolitis.[8-9] Using 16S rRNA sequencing, we found that these infants have altered fecal microbiota characterized by a very low diversity in gut microbial communities, which might make them more susceptible to developing necrotizing enterocolitis.

Q: Have any of your studies looked at how diet impacts the gut microbiome?
CT: We've performed several studies where we've used sequencing to identify diet-associated variations in the gut microbiomes of mice.[10-11] In a recent collaborative study with Drs. Hans-Rudolf Berthoud, Annadora Bruce-Keller, Michael Salbaum, and David Welsh, we performed an antibiotic knockdown of the microbial gut community in a group of mice that had been on a standard mouse chow diet. By oral gavage, we then transplanted in the microbiota from mice that had been fed either a high-fat diet (HFD) or a standard mouse chow diet. Sequencing-based phylogenetic analysis using the MiSeq System confirmed the presence of a very distinctive difference in microbiota between the groups. The mice given HFD microbiota also showed

Christopher Taylor, PhD is an Associate Professor at the Louisiana State University Health Sciences Center.

For Research Use Only. Not for use in diagnostic procedures.

significant and selective differences in laboratory measures of mouse behavior, such as fear conditioning, maze solving, and anxiety display. This data reinforced the link between gut dysbiosis and obesity-associated changes in neurocognitive behavior in mice.

> Sequencing-based phylogenetic analysis using the MiSeq System confirmed the presence of a very distinctive difference in microbiota between the groups.

Q: What other microbiome studies have you performed that enabled you to see the impact that microbiomes have on disease?
CT: We have conducted several genitourinary tract studies in collaboration with Drs. David Martin and Michael Ferris at LSUHSC. In one study, we looked at a disease called bacterial vaginosis and the role of *Gardnerella vaginalis*, which was originally thought to be the etiologic agent for that disease. A recent paper had shown that there are 2 fundamental *G. vaginalis* genotypes, a biofilm-forming pathogenic variant that was found in all patients with BV, and a commensal variant that does not form biofilms.[12]

In our study, we obtained samples from 53 women and their male sexual partners and used 16S rRNA sequencing to investigate the 2 *G. vaginalis* genotypes and their impact on disease.[13] During that investigation, a graduate student in my lab, Murat Eren, developed a new method called oligotyping that allowed us to differentiate between various *G. vaginalis* strains based only on 16S sequencing data.

Dr. Martin had observed a phenomenon in the clinic where women diagnosed with bacterial vaginosis would finish a course of antibiotics, and be cured, but would then return to the clinic several weeks later with the same symptoms. Our idea was that the bacteria were harbored by the male sexual partner and transmitted back to the woman during sexual activity following the completion of antibiotic treatment. When we looked at the bacterial 16S oligotyping profiles of *G. vaginalis* from the women and men in the study, we found that we could determine which participants were sexual partners, without a priori knowledge and with a high degree of certainty. The ability to look this deeply into strain level variation using only 16S rRNA sequencing was unprecedented and has since been extended with a method called Minimum Entropy Decomposition by Murat and his colleagues and applied to several interesting microbial community studies.[14]

Q: What is the value of 16S sequencing?
CT: 16S sequencing enables us to identify bacterial species present in a community and to separate similar strains using techniques like oligotyping. We're currently working on an airway study in collaboration with Dr. David Welsh where we're using 16S sequencing data to look at relatively subtle differences between microbiota sampled from the oral cavity with microbiota from tissue samples taken further down the airway with different sampling brushes. Compared to gut microbiota, the bacterial burden of airway microbiota is a lot lower. The further down the airway you go, the fewer bacteria you will find. We want to determine how these microbiota transit between these different environments and what this might tell us about the composition of microbiota living deeper in the airways.

Q: What types of environmental microbiome studies are you performing?
CT: We are planning a 16S sequencing study using the MiSeq System to analyze the metagenomic landscape of the Intensive Care Units (ICUs) in the new University Medical Center that is scheduled to open in mid-2015. We're very interested in looking at environmental samples from the new trauma ICU and medical ICU before patients are moved in and as these ICUs are put into service. Based on a trial run in another operating medical facility using qPCR to quantify 16S ribosomal RNA, we found the largest microbial counts came from the floor and from the lever on the hand sanitizer. In the new medical center, we're proposing to sample patient rooms and common areas where health care workers will be walking and using equipment, such as computer terminals, hand sanitizers, and scanners for tracking drug administration.

We also want to look at how antibiotic-resistant organisms end up entering and moving around in these new ICUs. One of our collaborators on the study is proposing a simultaneous culture-based investigation to look for particular antibiotic-resistant organisms. We're hoping to associate our 16S sequencing data with antibiotic-resistant and nonresistant versions of the cultured microorganisms and better understand how to control and prevent the spread of antibiotic-resistant microorganisms.

Q: What types of sequencing approaches did you use before obtaining a MiSeq System?
CT: In some of my initial research, we used genome tiling microarrays, but by the mid-2000s they had become obsolete with the advent of high-throughput sequencing. When I arrived at LSU, the School of Medicine had a pyrosequencing system that was being used to obtain long read lengths for 16S rRNA studies. This pyrosequencing system was older technology and was fraught with problems due to the large number of complex steps required for sample preparation. On many occasions, we struggled with this technology and had to perform multiple repeated sequencing runs attempting to resolve questionable results. It was also a long process, when you consider the time required for sample preparation, sequencing, and the weeks-long, intensive computation that was required to de-noise and prepare the data for analysis.

Q: How has the MiSeq System improved your workflow?
CT: We obtained a MiSeq System in August 2013, and the difference was like night and day. The cartridge-based MiSeq System eliminated many sources of variation where things could go wrong in sequencing preparation. Our sequencing throughput has been reduced from weeks to days. Due to the consistent and high-quality data generated by the MiSeq System, we were able to move forward with several of our projects. We didn't have to go back and repeat the same sequencing run over and over, trying to get usable data.

For our first MiSeq sequencing run, we used samples from the mouse gut repopulation study I mentioned earlier. When we ran these samples on the MiSeq System, the results were some of the best data we had ever seen. The beta diversity plot groupings obtained for the HFD mice versus the control diet mice were incredibly distinct and provided a better separation than we could have expected. We were

For Research Use Only. Not for use in diagnostic procedures.

even able to identify a mislabeled sample from the study due to the extremely high fidelity of the diet groupings.

> The cartridge-based MiSeq System eliminated many sources of variation where things could go wrong in sequencing preparation.

Q: What do you see for the future of metagenomics?

CT: One thing I envision is the prescriptive use of probiotics. Currently, there is a lot of pseudoscience in the supplement industry, promoting probiotics that are just different strains of *Lactobacillus*. The market for these products is based on the idea that *Lactobacillus* is generally helpful as a way to improve or maintain human health. Sequencing approaches will enable us to acquire a much more nuanced and specific understanding of what a healthy gut microbial community is and what kinds of treatments are possible, and the benefits of certain probiotics. Similar to the promise of personalized medicine, I think that in the future, we'll be able to determine a person's gut microbiota and the kinds of probiotics that could be most beneficial for them. That is something I see on the long-term horizon for microbiome research.

References

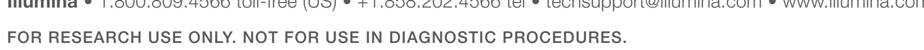
1. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004; 306(5696): 636–640.
2. Kamani N, Taylor C, Malhotra A, Dutta A. Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. Genome Res. 2007; 17(6): 865–876.
3. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447(7146): 799–816.
4. Lin Z, Puetter A, Coco J, et al. Detection of murine leukemia virus in the Epstein-Barr virus-positive human B-cell line, using a computational RNA-Seq-based exogenous agent detection pipeline, PARSES. J Virol. 2012; 86(6): 2970–2977.
5. Strong MJ, O'Grady T, Lin Z, et al. Epstein-Barr virus and human herpes virus 6 detection in a non-Hodgkin's diffuse large B-cell lymphoma cohort by using RNA sequencing. J Virol. 2013; 87(23): 13059–13062.
6. Xu G, Strong MJ, Lacey MR, Baribault C, Flemington EK, Taylor CM. RNA CoMPASS: a dual approach for pathogen and host transcriptome analysis of RNA-Seq data sets. PloS One. 2014; 9(2): e89445.
7. Strong MJ, Xu G, Morici L, et al. Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples. PloS Pathog. 2014; 10(11): e1004437.
8. Gupta RW, Tran L, Norori J, et al. Histamine-2 receptor blockers alter the fecal microbiota in premature infants. J Pediatr Gastroenterol Nutr. 2013; 56(4): 397–400.
9. McMurtry VE, Gupta RW, Tran L, et al. Bacterial diversity and *Clostridia* abundance decrease with increasing severity of necrotizing enterocolitis. Microbiome. 2015; 3:11 doi: 10.1186 /s40168-015-0075-8.
10. Wicks S. Taylor CM, Luo M, et al. Artemisia supplementation differentially affects the mucosal and lumenal ileal microbiota of diet-induced obese mice. Nutrition. 2014: 30(7-8 Suppl): S26-30. Doi: 10.1016/j.nut.2014.02.007.
11. Bruce-Keller AJ, Salbaum JM, Luo M, et al. Obese-type Gut Microbiota Induce Neurobehavioral Changes in the Absence of Obesity. Biol Psychiatry. 2015; 77(7): 607–615.
12. Harwich MD Jr, Alves JM, Buck GA, et al. Drawing the line between commensal and pathogenic *Gardnerella vaginalis* through genome analysis and virulence studies. BMC Genomics. 2010; 11:375 doi: 10.1186/1471-2164-11-375.
13. Eren AM, Zosaya M, Taylor CM, Dowd SE, Martin DH, Ferris MJ. Exploring the diversity of *Gardnerella vaginalis* in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. PloS One. 2011; 6(10): e26732.
14. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J. 2015; 9:968–979.

Learn more about the Illumina system mentioned in this article:

- MiSeq System, www.illumina.com/systems/miseq.html

For Research Use Only. Not for use in diagnostic procedures.

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



FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

© 2015 Illumina, Inc. All rights reserved. Illumina, MiSeq, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. Pub. No. 1270-2015-002 Current as of 14 May 2015

Wine, Cheese, and the Microbiomes That Influence Their Flavor

Researchers use the MiSeq® System to identify the microbiome profiles of wine and cheese, and the facilities in which they are made.

Introduction

Most microbiology studies focus on how to identify and remove microbes from our food—and for good reason. Pathogens like *Salmonella enterica*, *Listeria monocytogenes*, and *Escherichia coli* sicken countless people every year. But University of California, Davis (UC Davis) food scientist David Mills, PhD, is quick to point out that microbes are actually the true chefs behind some of our favorite foods. Lactic acid bacteria metabolize the lactose in milk, helping convert it into the savory cheeses that we enjoy such as cheddar, brie, or pecorino romano. And the yeast *Saccharomyces cerevisiae* performs its magic on grapes, transforming their juice into a seemingly infinite variety of wines for every taste and occasion.

Dr. Mills studies the microbiomes of cheese and wine, and the facilities that produce them, to understand how microbes work their magic. It's a question that scientists have only recently been able to tackle, thanks to next-generation sequencing (NGS) instruments, like the MiSeq System.

He hopes to discover how cheeses and wines cultivated in different regions develop their own distinctive flavors by observing and cataloging their unique constellation of microbes. "The concept of regionality in food might be traceable through the microbes that are associated with it," Dr. Mills said. "The microbiota profile at the front end, before the grapes are fermented, could correlate with the profile of the chemical constituents of the finished wine. A very exciting aspect of our research is that we might be able to predict flavor outcomes by studying these microbes."

A Different Wine and Cheese Pairing

Dr. Mills began his career in biochemistry and biotechnology. As a PhD student at the University of Minnesota, he switched to the microbiology program after becoming interested in the work of Larry McKay, PhD, a microbiologist who studies the bacteria that make cheese. This appealed to Dr. Mills' interest in translational science. "While I enjoy basic research, what drives me is studying its application in processes and the environment," Dr. Mills stated.

He decided to focus on the genetics of *Lactococcus lactis*, one of the lactic acid bacteria needed to make cheddar cheese. After a short postdoc at North Carolina State University, Dr. Mills was hired by the renowned Department of Viticulture and Enology at UC Davis. Although wine's main fermenter is *S. cerevisiae*, lactic acid bacteria also perform some of the fermentation. It was a wine and cheese pairing of a different sort, and Dr. Mills loved it.

Sampling Cheesemaking Facility Microbiomes

Understanding the microbial ecology of any food involves more than a focused study of the hundreds or thousands of different microbes found in the food itself. It also means studying the microbes found in the facility where the food is prepared and packaged.

"We're very interested in understanding the microbial landscape of the whole food production facility, such as the seasonality of microbial movement inside the facility," Dr. Mills said. "For example, how activities inside the facility lead to microbial transfer back and forth.

David Mills, PhD, is a Professor in the Departments of Food Science and Technology, and Viticulture and Enology at the UC Davis.

Photo courtesy of UC Davis Communications

UC Davis viticulture and enology student Javier Avina prepares a fermenter containing crushed grapes for primary fermentation.

For Research Use Only. Not for use in diagnostic procedures.

Often, I'm studying facilities where microbes are purposely introduced for making cheese or wine. The question is, do these microbial colonies grow to populate the entire place? Where else do they end up? In instances where sanitation is performed correctly, we want to understand how other microbes present in the facility take over."

No matter the type of facility being studied, identifying these microbes means taking samples—lots of them. Dr. Mills and colleagues use cotton swabs to take thousands of samples from buildings and the foods they produce.

After chemically isolating and extracting the DNA, Dr. Mills originally employed "very rudimentary tools" to identify the different microbes by hand, using denaturing gradient gel electrophoresis. "Old school methods had very low throughput," Dr. Mills said. "We could only analyze 20 samples at a time and the resolution was terrible. Graduate students hated it because it was a tough method and very time-intensive. We could get an indication that there were a bunch of microbes present, but we weren't sure what they were."

Role of House Microbiota in Cheese and Winemaking

While the interest in decoding the microbes that make wine, cheese, and beer has always been high, the research stalled in the absence of efficient tools. This changed several years ago, when the sequencing core at UC Davis invested in NGS and purchased several MiSeq Systems. Taking advantage of this advanced technology, Dr. Mills began identifying microbes from large numbers of samples. Together with graduate student Nickolas Bokulich, he began to look more closely at the microbes in a California artisanal creamery to see if a facility-specific or "house microbiota" could be identified. They sampled microbes from the surfaces of the washed rind cheeses (a class of cheese that includes Limburger and other pungent varieties), and examined some of the surfaces in the facility itself.[1]

"We could see that the microorganisms on the surface of a cheese were the same ones present on the surface of some parts of the facility," Dr. Mills stated. "We're still unsure which direction the microbes colonized: from the cheese to the facility surfaces, from the surfaces to the cheese, or in both directions over time."

Regardless of the colonization dynamics, the results supported the idea that a house microbiota helps to inoculate a cheese or wine and cause it to take on a specific flavor. Vintners in the Bordeaux region of France have said for years that the winery itself inoculates the wine. It's a concept called *terroir*, the idea that the place where a wine or food is grown or made is ultimately expressed in its final taste.

"People will taste a wine and say 'that's a Bordeaux,'" Dr. Mills said. "A good sommelier can even tell you the vintage and region of origin. Yet, we don't know what causes that regionality. How do we connect the sense of place to actual sensory metabolites in the wine, so we can understand why a Napa Valley cabernet delivers a specific flavor that can't be found in a cabernet from the California Central Valley?"

No one knows exactly what creates *terroir* in a food, but Dr. Mills' work is showing that microbes likely play a role. His sequencing analysis studies begin with extracting DNA from samples and amplifying portions of bacterial and fungal ribosomal genes with different sets of barcoded primers, some of which were developed in his lab. The library sets are sent to the UC Davis Genome Center, where the MiSeq System is used to sequence 500 samples in a single run. The amplicon sequences tell the researchers which microbial species are present in a given sample, allowing them to take a census of microbes across a facility and over time.

Identifying Winery Microbiomes

Back at the UC Davis winery, Dr. Mills and Bokulich sampled the whole facility about 150 times using cotton swabs at several points throughout the winemaking season. They started in August, before the grapes were harvested and brought inside. They returned in the middle of the season, as the grapes were brought into the winery to be crushed. And they made a final visit in mid-December, when the facility was being cleaned and the winemaking equipment stored for the season. As they predicted, they saw dramatic shifts in the microbes over that time. They also found some things they didn't expect to see. For example, *S. cerevisiae* arrived on the doorstep of the winery long before a single grape came in, perhaps brought in by workers walking back and forth from the vineyard to the winery.[2]

Dr. Mills and his colleagues expanded their study to wineries throughout California, looking at the microbes that grew on grapes and in the facilities over 1 to 2 seasons. The goal was to see if microbial populations were unique for different wines from different regions, and to study how these microbes moved through the system from harvesting to the final product. They also sought to document the role of microbes in producing a specific wine and its flavor.

Clearly, the flavor molecules in the grapes themselves strongly influence the final flavor of the wine, and certain regions produce grapes that yield better wines than others. "Whether microbes play a 1% or an 80% part of the flavor-generation process, I don't know," Dr. Mills stated. "The Holy Grail in wine is to understand where a good wine flavor comes from, and we know that there is an amalgam of microbes needed to produce the final product. Many of the most-desired wines have much more complex flavors that are driven by more than just *S. cerevisiae*. We're attempting to understand the constellation of microbes that delivers that actual flavor."

The research showed that "there are different sets of microbes entering the wineries from different regions on the same grapes," Dr. Mills said. "We know some of those microbes influence the flavor of wine, providing the potential connection between something in the vineyard to a flavor molecule in the wine."

Dr. Mills and his team studied *must*, the freshly pressed juice created from crushed grapes before they are fermented.[3] The data revealed that the microbes entering a winery on *grape must* are region-specific. For example, the microbes on *chardonnay must* from the California Central Coast are different from those on *chardonnay must* from Napa Valley or Sonoma County.

"If you take the same vine type and grow it in Sonoma and in Napa, I wouldn't expect to see that much difference in the microbial landscape," Dr. Mills stated. "Plants, like humans, have typical microbes associated with them. So I was a bit surprised at how robustly we could differentiate the regions based on the microbial communities we found."

Different populations of microbes that enter into fermentation could cause flavor differences in the final product, a hypothesis that Dr. Mills and other scientists are just beginning to test.

For Research Use Only. Not for use in diagnostic procedures.

NGS Enables Efficient, High-Quality Microbiome Research

None of this work would have been possible without NGS systems, such as the MiSeq System, and bioinformatics programs that help researchers make sense of the data. “The irony of the situation is that collecting the samples is now the biggest problem I have in performing these studies,” Dr. Mills said. “Processing the samples and running them on a sequencer is easier than going to a facility and spending 8 hours obtaining 300 swabs, all documented so we know exactly where the sample was taken. In one of the papers we’re submitting soon, the winery had to collect about 2000 fermentation samples for us. That’s a large number of samples and they all have to be labeled correctly and stored somewhere. It’s a lot of work.”

For Dr. Mills’ research, the MiSeq System provides an optimum combination of speed and cost. “I like the MiSeq System because it’s a scaled down version of a high-throughput sequencer,” Dr. Mills stated. “It turns around samples quickly and I know that the quality is good. While I’d love to see a larger read, I’m happy with the reads that I get.”

His work has implications beyond identifying the microscopic creatures that help to make some of our favorite foods and wine. It might also help us understand how microbes move through our world.

“Our goal is to develop a microbial map of a whole facility,” Dr. Mills added. “One of the beauties of using NGS tools is that we can process a large number of samples. Our MiSeq runs are often 500 samples or more. In the old days, plating 500 samples would not have been feasible. We could never have performed the kind of high-throughput studies that we can today. With NGS data, we are opening up a completely new window, enabling us to view the microbial landscape of food facilities.”

References

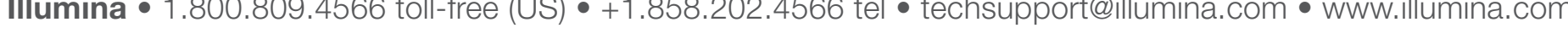
1. Bokulich NA and Mills DA. Facility-specific “house” microbiome drives microbial landscapes of artisan cheesemaking plants. Appl Environ Microbiol. 2013;79(17): 5214–5123.
2. Bokulich NA, Ohta M, Richardson PM, Mills DA. Monitoring seasonal changes in winery-resident microbiota. PloS One. 2013;8(6):e66437.
3. Bokulich NA, Thorngate JH, Richardson PM, Mills DA. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. Proc Natl Acad Sci USA. 2014;111(1):139–148.

Learn more about the Illumina system mentioned in this article and the art of cheesemaking:

- MiSeq System, www.illumina.com/systems/miseq.html
- Illumina SciMon Video: Cheesemaking: Ancient Art or the Simplicity of Science, www.youtube.com/watch?v=M0QUzpRCvP4&feature=youtu.be&list=PLKRu7cmBQlahAeoVCcb-pM1oVCKFkggFh

Follow the Illumina Agrigenomics Community

For Research Use Only. Not for use in diagnostic procedures.



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

For Research Use Only. Not for use in diagnostic procedures.

© 2015 Illumina, Inc. All rights reserved. Illumina, MiSeq, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. Pub. No. 1370-2015-001 Current as of 14 May 2015

Metagenomic Analysis of Environmental Water Samples With the NextSeq® 500 System

Shotgun metagenomic sequencing provides insight into microbial responses to environmental changes in a water reservoir.

Introduction

Environmental metagenomics is the study of organisms in a microbial community by analyzing the DNA present in an environmental sample. The advent of next-generation sequencing (NGS) technology has enabled researchers to profile entire microbial communities in complex samples quickly and easily. Unlike capillary electrophoresis or PCR, NGS allows investigators to sequence thousands of organisms in parallel. With shotgun metagenomic sequencing, microbiologists can examine the genes of the organisms in a given sample comprehensively to evaluate bacterial diversity and detect microbial abundance in various environments. When designing a shotgun metagenomic sequencing study, researchers must consider many factors, including the complexity of the sample and the sequencing output required to assess microbial diversity. Illumina library preparation kits and the NextSeq Series of sequencing systems delivers the efficiency, high throughput, speed, and sample size flexibility needed for efficient and affordable metagenomics studies.

This application note describes how investigators from the Center for Earth and Environmental Science and Indiana University-Purdue University Indianapolis (IUPUI) used shotgun metagenomic sequencing to characterize microbial communities in samples from the Eagle Creek reservoir in Indianapolis, Indiana. The method includes sample collection, DNA extraction using the Meta-G-Nome™ DNA Isolation Kit,[1] DNA library preparation with the Nextera® XT DNA Library Prep Kit,[2] sequencing with the NextSeq 500 System,[3] and analysis using the MG-RAST metagenomics analysis server[4] (Figure 1). The results presented in this study provide new insights into the taxonomic and functional differences between microbial communities in the Eagle Creek drinking water supply reservoir at different times and locations.

Methods

Sample Collection

In collaboration with Citizens Water Company,[5] the Center for Earth and Environmental Science (IUPUI) regularly samples, monitors, and documents the water supply in the Eagle Creek reservoir. The samples for this study were collected at regular intervals, from May to October 2013, before the first frost. Discrete samples were collected from the surface, and at the depths of 3 meters, 6 meters, and near the bottom floor of the reservoir. During the sampling process, 150 ml of lake water were collected, filtered through a 0.22 µm filter, and frozen at -20°C.

Genomic DNA Isolation

After sample collection, DNA was isolated directly from the water sample using the Meta-G-Nome DNA Isolation Kit (Epicentre®, an Illumina company). The protocol uses filtration technology and enzymatic lysis to isolate DNA from the water sample. The kit is designed to isolate randomly sheared genomic DNA of high molecular weight.

Library Preparation

The Nextera XT DNA Library Prep Kit was used to construct libraries from the isolated DNA. Using a single enzymatic "tagmentation" reaction, the Nextera transposome simultaneously fragmented and tagged the DNA with unique adapter sequences. Limited-cycle PCR was used to amplify the tagged DNA and add sequencing indexes. Using this streamlined workflow, 11 DNA libraries were prepared for sequencing on the NextSeq 500 desktop sequencer in half a day.

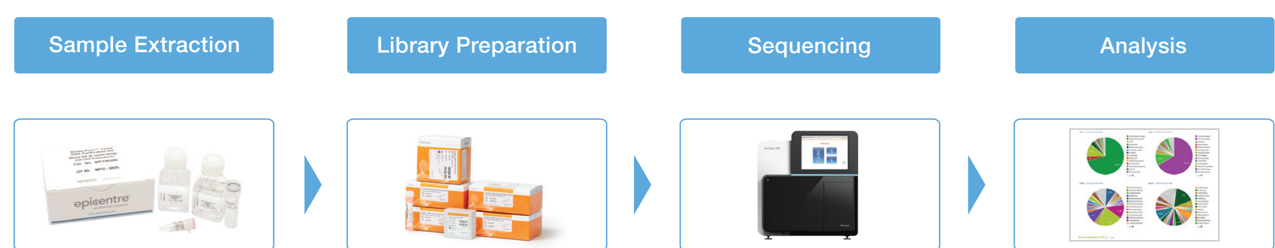


Figure 1: Shotgun Metagenomic Sequencing Workflow—The speed of the Nextera XT DNA Library Prep Kit and the output of the NextSeq Series contribute to a fast and efficient workflow suited to NGS-based environmental metagenomics studies.

For Research Use Only. Not for use in diagnostic procedures.

Sequencing

All 11 libraries were pooled together for cluster generation and sequencing. Libraries were loaded onto a reagent cartridge and clustered on the NextSeq 500 System. A paired-end, 2 × 150 bp sequencing run was performed using the NextSeq 500 High-Output Kit. A single sequencing run generated 400 million reads in 29 hours, corresponding to an average of 40 million reads per sample after quality filtering (Table 1). Base calls generated by the NextSeq 500 System were converted to FASTQ files for metagenomic analysis.

Data Analysis

Data analysis is often the most challenging part of a metagenomic sequencing study. Several programs and pipelines are available to perform metagenomic analysis. These programs are optimized for different study objectives, such as taxonomic profiling, assessing microbial composition, or identifying functional genes and pathways.

The analysis presented in this study used the publicly available MG-RAST program, an automated analysis platform that provides various tools for data visualization enabling users to assess species composition or functional abundance. FASTQ files generated by the NextSeq 500 System and sample metadata (Table 2) were uploaded to the MG-RAST server. Overlapping paired-end reads were joined and then processed in MG-RAST using the default trimming and filtering settings to allow comparison to other data sets in MG-RAST.

Table 1: Sequencing Data

MG-RAST ID	Metagenome Name	Base Count[a]	Sequence Count[b]
4554374.3	1305-530	5,061,868,863	33,395,202
4554375.3	1305-531	5,566,909,825	37,789,429
4554376.3	1305-532	8,409,674,885	57,196,837
4554377.3	1305-533	4,341,474,376	31,003,711
4554378.3	1307-521	5,184,242,473	32,673,383
4554379.3	1307-522	14,173,422,835	85,796,130
4554380.3	1307-523	8,369,469,023	52,993,984
4554381.3	1307-524	2,864,009,160	19,306,387
4554382.3	1310-522	8,688,763,557	55,503,302
4554383.3	1310-523	10,943,344,357	69,814,786
4554384.3	1310-524	7,776,089,825	46,838,779

a. All sequenced bases
b. Number of sequencing reads

Table 2: Environmental Metadata

MG-RAST ID	Sampling Depth	Sampling Date	Water Temperature (°C)	pH	Air Temperature (°C)	Location Coordinates[a]	Salinity (ppm)
4554374.3	Surface	23 May 2013	21.05	8.72	13.3	39.827, -86.303	0.21
4554375.3	3 meters	23 May 2013	21.00	8.73	13.3	39.827, -86.303	0.21
4554376.3	6 meters	23 May 2013	13.77	8.27	13.3	39.827, -86.303	0.22
4554377.3	Floor	23 May 2013	10.78	7.50	13.3	39.827, -86.303	0.20
4554378.3	Surface	25 July 2013	26.50	8.48	18.9	39.826, -86.304	0.21
4554379.3	3 meters	25 July 2013	26.02	8.40	18.9	39.826, -86.304	0.23
4554380.3	6 meters	25 July 2013	23.03	7.60	18.9	39.826, -86.304	0.22
4554381.3	Floor	25 July 2013	13.86	7.16	18.9	39.826, -86.304	0.23
4554382.3	3 meters	23 October 2013	15.09	7.80	4.4	39.826, -86.304	0.24
4554383.3	6 meters	23 October 2013	15.06	7.72	4.4	39.826, -86.304	0.24
4554384.3	Floor	23 October 2013	14.99	7.68	4.4	39.826, -86.304	0.24

a. Global Positioning System (GPS) coordinates

For Research Use Only. Not for use in diagnostic procedures.

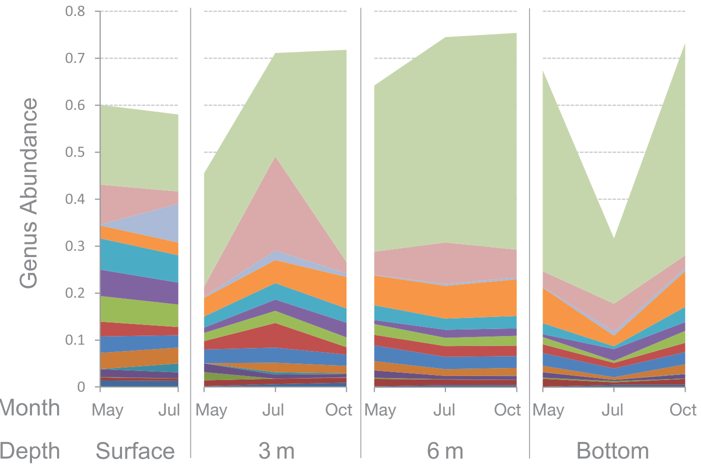


Figure 2: Comparative Genus Abundance Over Time—Comparisons of relative genus abundance (shown for the 15 most abundant genera) demonstrated a sharp decline in *Arthrobacter* populations on the reservoir floor in July 2013. These data were generated using MG-RAST.

Results

Analysis of species abundance with MG-RAST revealed drastic changes in microbial composition on the reservoir floor and an increase in the *Rhodococcus* genus (a soil bacteria), at the 3 meter depth in July 2013 (Figure 2). For the reservoir floor, the population declines might be correlated with an algaecide treatment that occurred on 31 May 2013. It is possible that the copper in the algaecide treatment caused the sharp decline in species from the *Arthrobacter* genus. This hypothesis is based on the observation that *Arthrobacter globiformis* often forms a symbiotic relationship with the *Anabaena* genus of nitrogen-fixing cyanobacteria and the aquatic ferns in the *Azolla* genus, both part of the algae family.[6] After algaecide treatment, both *Arthrobacter globiformis* and *Azolla* experienced a decline; however, *Azolla* is not shown in Figure 2, because it is not 1 of the 15 most abundant genera in the samples. The hypothesis that the gram-positive *Anabaena* species might have played a role in the changes to community composition requires further analysis. Likewise for the 3 meter sampling depth, where analysis revealed an increase in the *Rhodococcus* genus, further analysis is required to assess root cause.

Conclusions

This study demonstrates how shotgun metagenomic sequencing with the NextSeq 500 System can reveal information about the microbial composition of a particular environment. The data presented show a correlation between the environmental metadata and the species composition in the Eagle Creek reservoir. This study provides new insights into the biological processes potentially associated with algal blooms, sampling depth, and seasonal differences in freshwater environments.

Learn More

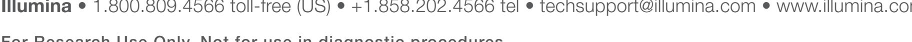
To learn more about the NextSeq Series, visit www.illumina.com/nextseq.

For more information about the use of Illumina technology in microbial genomics, visit www.illumina.com/microbiology.

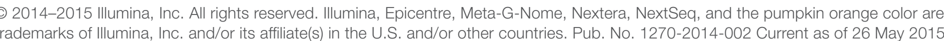
References

1. Meta-G-Nome- DNA Isolation Kit (www.epibio.com/applications/nucleic-acid-purification-extraction-kits/dna-purification-genomic/meta-g-nome-dna-isolation-kit) Accessed 21 April 2014.
2. Nextera XT DNA Library Prep Kit (www.illumina.com/products/nextera_xt_dna_sample_prep_kit.ilmn) Accessed 21 April 2014.
3. NextSeq 500 Desktop Sequencer (www.illumina.com/systems/nextseq-sequencer.html) Accessed 21 May 2015.
4. Meyer F, Paarmann D, D'Souza M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics.* 2008;9:386.
5. Citizens Energy Group (www.citizensenergygroup.com) Accessed 21 April 2014.
6. Martens T, Gram L, Grossart HP, et al. Bacteria of the *Roseobacter* clade show potential for secondary metabolite production. *Microb Ecol.* 2007;54:31-42.

For Research Use Only. Not for use in diagnostic procedures.


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

For Research Use Only. Not for use in diagnostic procedures.


© 2014–2015 Illumina, Inc. All rights reserved. Illumina, Epicentre, Meta-G-Nome, Nextera, NextSeq, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. Pub. No. 1270-2014-002 Current as of 26 May 2015