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.

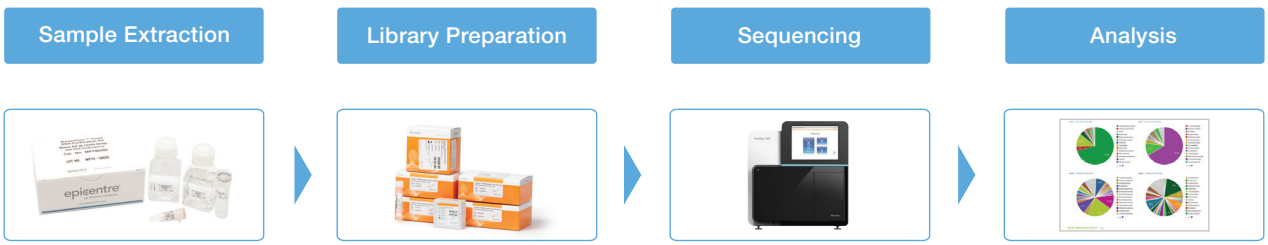

Figure 1: Shotgun Metagenomic Sequencing Workflow—The speed of the Nextera XT DNA Library Prep Kit and the output of the NextSeq Series contribute to a fast and efficient workflow suited to NGS-based environmental metagenomics studies.

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.

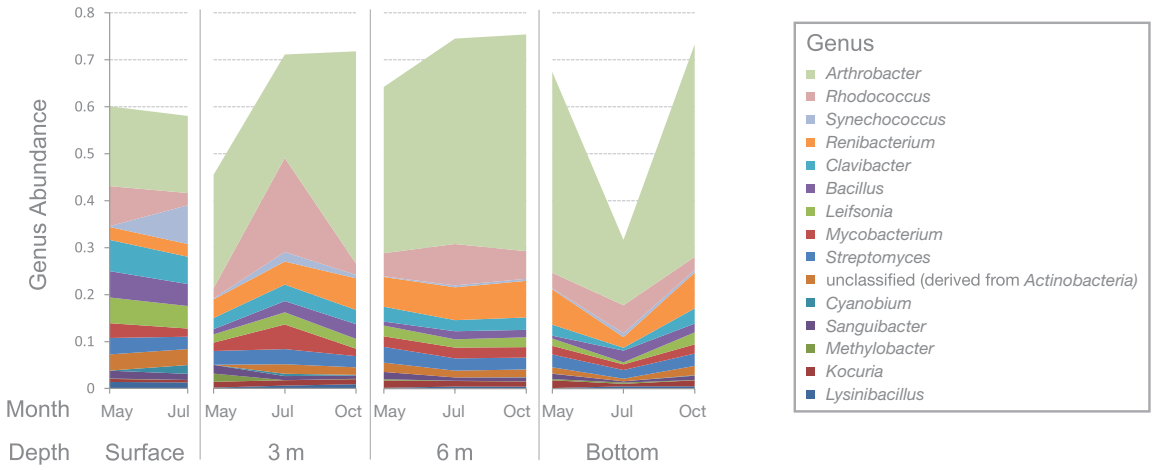

Figure 2: Comparative Genus Abundance Over Time—Comparisons of relative genus abundance (shown for the 15 most abundant genera) demonstrated a sharp decline in *Arthrobacter* populations on the reservoir floor in July 2013. These data were generated using MG-RAST.

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.

AAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAA
AATCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACC
AACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTA
TTAACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGAT
AAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAA
AAGATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGA
AACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAG

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

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA