Nextera™ DNA Flex Library Preparation for Soil Shotgun Metagenomics Analysis

Explore taxonomic and functional diversity of soil microbial communities with a comprehensive shotgun metagenomics sequencing workflow.

Introduction

During the last decade, the human microbiota has been an area of immense interest with regards to its impact on human health. With advancements in next-generation sequencing (NGS) technologies, profiling prokaryotic microbes originating from human samples has been accomplished predominantly through 16S sequencing, an amplicon-based approach. Resulting data support an integral role for bacterial communities in maintaining human health. However, 16S sequencing does not reveal the entire microbial community and typically provides genus-level taxonomic assignments, limiting the amount of information that can be learned from the data.Conversely, bacterial communities can promote various diseases, such as Crohn's disease, which may result from imbalances in gut bacterial composition.[1]

The soil microbiome represents a highly diverse and complex microbial community that contributes to many aspects of human, animal, and environmental health. Widely regarded as one of the most diverse habitats on Earth, the soil microbiome is composed of taxa from the four main domains of life: Bacteria, Archaea, Eukarya and viruses.[2] Together, these microbes play a dynamic role in influencing our ecosystem through nutrient cycling, organic matter decomposition, antibiotic development/resistance, and more.[3] Remarkably, the majority of soil microbes have yet to be isolated and their functions are largely unknown. Environmental perturbations such as changes in pH, moisture, and sodium content can drastically change microbial diversity and functions.[3] Using NGS workflows optimized for culture-free sample preparation and sequencing enables unbiased profiling of soil microbe composition. Shotgun sequencing of soil-derived genomic DNA (gDNA) is an unbiased approach for metagenomic profiling that enables species-level taxonomic classification, *de novo* assembly, genomic functional profiling, and more.

The Nextera DNA Flex Library Preparation Workflow for Soil Metagenomics is a comprehensive solution for scalable soil metagenomics studies. This application note demonstrates a robust workflow for soil gDNA extraction, streamlined library preparation, shotgun sequencing, and simplified metagenomic data analysis apps in BaseSpace™ Sequence Hub (Figure 1).

Methods

To validate this workflow, stool samples were analyzed and used as a control to demonstrate the significantly greater diversity of soil compared to stool.

Soil collection

Soil samples were collected from various sites and geographic regions across the United States (Table 1). Samples were stored at 4°C in 50 mL conical tubes until DNA isolation.

Table 1: Soil samples

Sample type	Region	Sample type	Region
Oak canopy	Escondido, CA	Riparian	Estes Park, CO
Escondido creek	Escondido, CA	Wildflower	Estes Park, CO
Riparian	Encinitas, CA	Corn field	Madison, WI
Dry coastal scrub	Encinitas, CA	Wheat field	Madison, WI
Wild oat	Encinitas, CA	Garden	Madison, WI
Near creek	Encinitas, CA	Soybean field	Madison, WI

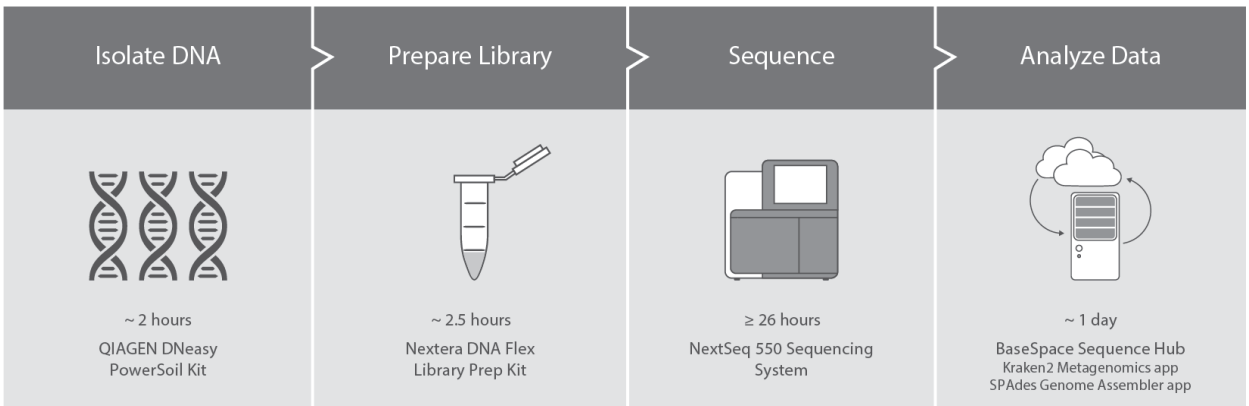

Figure 1: Nextera DNA Flex Soil metagenomics NGS workflow—Nextera DNA Flex Library Preparation is part of a comprehenisve NGS workflow for soil metagenomics analysis.

DNA extraction: soil

Extractions were performed manually using the DNeasy PowerSoil Kit (catalog no. 12888-100, QIAGEN).[4] Each extraction was performed with 0.25 mg of soil input and eluted with a volume of 100 µl. This extraction method robustly produces ≥ 1 ng/µL DNA, which is within range for the Nextera™ DNA Flex Library Prep kit input concentration. Integrity and concentration of the DNA was assessed with the Fragment Analyzer (catalog no. M5311AA, Agilent Technologies) using the HS NGS Fragment kit (catalog no. DNF-474-1000, Agilent Technologies).[5]
This method for isolating gDNA from soil is ideal as it eliminates PCR and library preparation inhibitors, such as humic substances, enabling detection of a wide variety of microbes.

DNA extraction: stool

Stool samples were provided by PerkinElmer, extracted using the chemagic DNA Stool 200 Kit special H96 (Catalog No. CMG-1076, PerkinElmer) and the automated chemagic 360 instrument (catalog no. 2024-0020, PerkinElmer).

Library preparation

Sixteen Nextera DNA Flex libraries (12 soil samples, 4 stool samples) were prepared manually using the Nextera DNA Flex Library Prep Kit (Catalog No. 20018705, Illumina).[6] The total DNA input (100–400 ng) was normalized to a volume of 30 µl with nuclease-free water prior to on-bead tagmentation. The quality and concentration of PCR-amplified libraries were assessed using the Fragment Analyzer prior to pooling.

Sequencing

Pooled libraries were sequenced on the NextSeq™ 550 System with a run configuration of 2 × 150 bp. For larger studies, sequencing runs can be scaled up to the NovaSeq™ 6000 System (S2 flow cell).

Data analysis

Pooled libraries were demultiplexed in BaseSpace Sequence Hub, the Illumina genomics computing platform. Taxonomic classification and down-sampling were conducted through the Kraken2 Metagenomics and FASTQ Toolkit Apps, respectively.[7,8]
Alpha diversity measurements and sample clustering were performed with analytical tools provided by MetaHit Consortium, using Kraken2 output data.[9] Prior to gene identification, metagenomes were assembled using SPAdes Genome Assembler, also available through BaseSpace Sequence Hub (Table 2).[10]
Gene detection was performed using online comparative analysis tools available through Joint Genome Institute (JGI) (GOLD for project submission and IMG/M for gene detection and functional profiling).[11,12]

Table 2: Metagenomics sequencing BaseSpace apps

BaseSpace App	Description
 Kraken2 Metagenomics	The Kraken2 Metagenomics app assigns taxonomic labels to short DNA sequences with high sensitivity and speed.
 SPAdes Genome Assembler	The SPAdes Genome Assembler is designed to assemble small genomes from standard bacterial data sets.
 FASTQ Toolkit	The FASTQ Toolkit provides a modular set of analyses for quality control checks on raw sequence data before downstream analysis.

Results

Taxonomic classification of soil samples

A total of 12 soil samples were collected from various regions of California, Colorado, and Wisconsin to assess the ability to use Nextera DNA Flex shotgun sequencing for soil metagenomics analysis. To gain an understanding of the microbial composition for each sample, taxonomic classification was performed with Kraken2 and data was visualized using Pavian[13] (Figure 2). Despite limitations with availabile microbial reference genomes, one of the advantages of shotgun metagenomic sequencing over 16S sequencing is the high resolution of taxonomic identification, allowing for robust species-level detection.

By subsampling three million classified reads from each soil sample, more than 1000 prokaryotic species were detected. Many of these detected species have known functions within the soil microbial community, such as nitrogen fixation.[14,15] Taxa of relatively high abundance, such as *Rhodobacter sphaeroides* (Bacteria) and *Bradyrhizobium genera*, are known to have symbiotic relationships with plants to fix atmospheric nitrogen (Figure 3).

Detection of nonprokaryotic microbes

The significant diversity of the soil microbiome has been well established. Many of these taxa belong to the Eukarya and virus domains, such as fungi and a broad range of bacteriophages, respectively, and are known to influence plant growth and microbial functions.[16,17] The ability to detect nonprokaryotic microbes using shotgun metagenomics sequencing provides an additional layer of information for metagenomic profiling, allowing for deeper characterization of specific microbial communities (Figure 4).

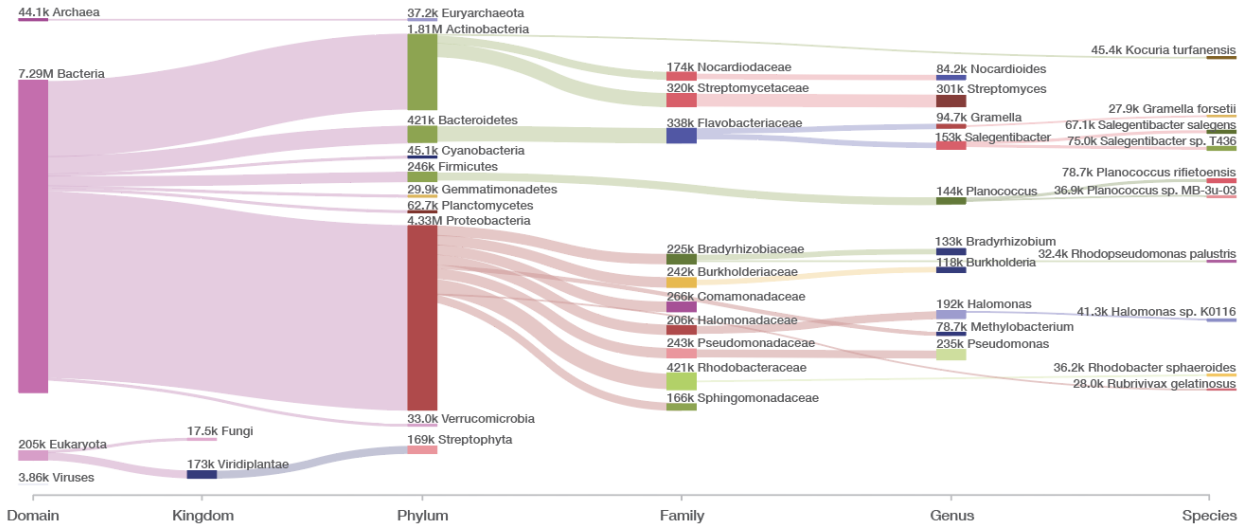

Figure 2: Taxonomic classification—Shotgun metagenomics sequencing enables high-resolution taxonomic classification and robust species-level detection.

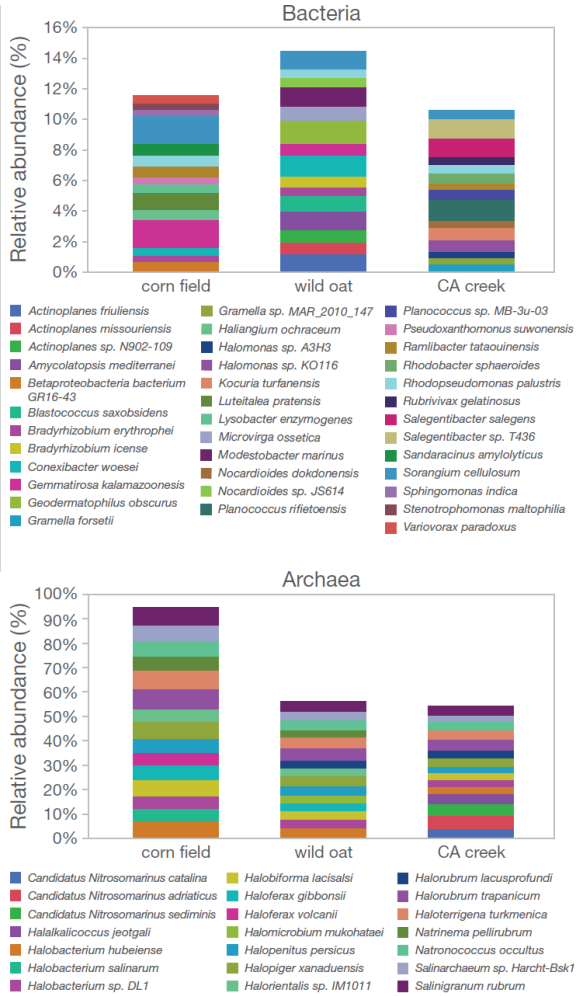

Figure 3: Relative abundance of top microbial species—The top 15 bacterial (top panel) and Archaea (bottom panel) species detected for the indicated soil samples are displayed with relative abundance.

Figure 4: Taxonomic classification of nonprokaryotic microbial species—Shotgun metagenomics sequencing enables taxonomic classification of eukaryotic and viral species within the Escondido creek soil sample.

Comparison of diversity in soil and stool samples

To determine the complexity of each soil microbial community, the richness and Shannon index were calculated to measure the number of species detected and overall diversity, respectively. The alpha diversity metrics demonstrated that the microbial diversity of soil is far greater than stool, consistent with current research findings (Figure 5).[18] Subsampling of sequencing data sets also revealed that over 10 million classified reads are necessary for a robust measurement of the number of detectable species from each soil sample (Figure 6).

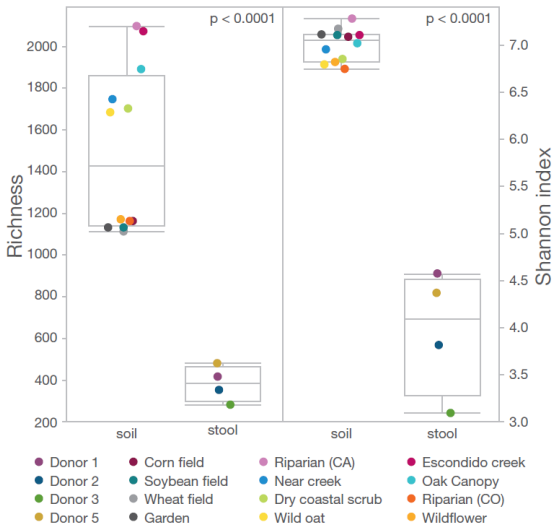

Figure 5: Higher microbial diversity in soil vs. stool samples—Measurement of diversity of each sample by richness (left) and Shannon index (right) indicate greater diversity of soil samples than stool.

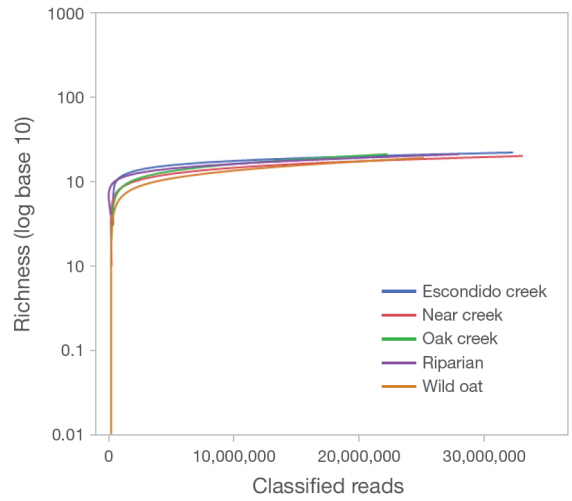

Figure 6: Comprehensive assessment of microbial composition requires > 10 million classified reads.

Clustering of samples by type and region

The performance of the Nextera DNA Flex Library Preparation Kit was further evaluated with heat mapping to cluster stool and soil samples. A specific signature was identified that clustered soil samples distinctly from stool samples, where a subset of highly represented species influenced this partition (Figure 7). Moreover, hierarchical clustering not only separated soil from stool samples, but it clustered soil samples by region of collection, further supporting the usage of Nextera DNA Flex for metagenomic profiling (Figure 8).

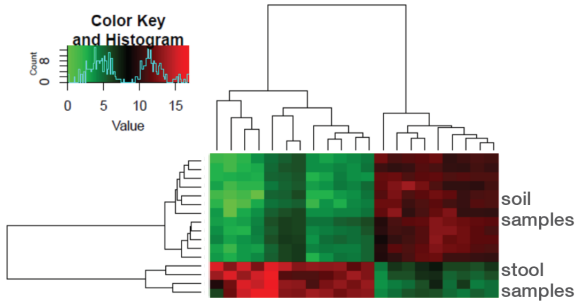

Figure 7: Distinguishing sample types by represented species—Heat map clustering reveals distinct signatures for soil and stool samples, with highly represented species influencing the division.

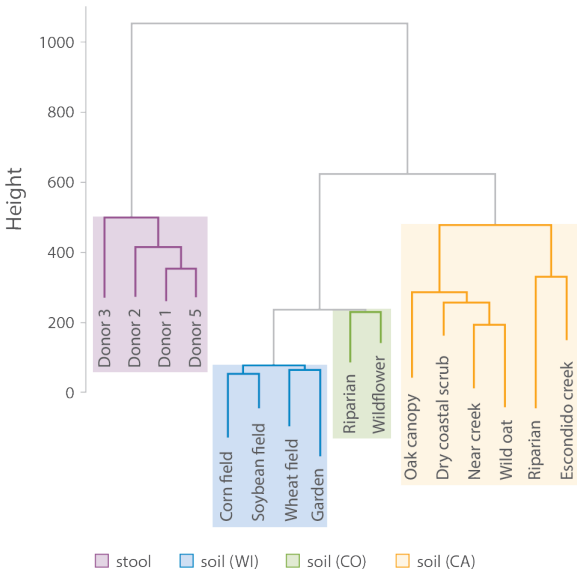

Figure 8: Hierarchical clustering of samples by type and region—Shotgun metagenomics sequencing of Nextera DNA Flex libraries enabled hierarchical clustering by sample type and region.

Insights into cellular function with shotgun metagenomics sequencing

To determine the functional potential of each soil microbial community, a gene detection algorithm was employed on assembled soil metagenomes via JGI.[19] Over 4 million genes were robustly detected for each metagenome using ~ 48 million paired

Table 3: Gene detection of soil metagenomes

Category	Counts[a]	
	Wild oat	Escondido creek
Genome size (nucleotides)	1,180,699,549	1,202,393,462
Protein-coding genes	4,373,522	4,341,193
rRNA genes	2058	2814
tRNA genes	6406	7568

a. ~48 million paired end reads were used for assembly and gene annotation.

Table 4: Functional profiling of soil metagenomes

Function ID	Gene name	Wild oat	Escondido creek
KO:K02004	putative ABC transport system permease protein (ABC.CD.P)	8656	8794
KO:K08884	serine/threonine protein kinase, bacterial [EC:2.7.11.1] (K08884)	12608	7836
KO:K03088	RNA polymerase sigma-70 factor, ECF subfamily (SIG3.2, rpoE)	9571	6551
KO:K02014	iron complex outermembrane recepter protein (TC.FEV.OM)	5438	6427
KO:K06147	ATP-binding cassette, subfamily B, bacterial (ABCB-BAC)	6038	5385
KO:K01990	ABC-2 type transport system ATP-binding protein (ABC-2.A)	7062	5333
KO:K03657	DNA helicase II / ATP-dependent DNA helicase PcrA [EC:3.6.4.12] (uvrD, pcrA)	3161	3035
KO:K02337	DNA polymerase III subunit alpha [EC:2.7.7.7] (DPO3A1, dnaE)	2160	2937
KO:K07712	two-component system, NtrC family, nitrogen regulation response regulator GlnG (glnG, ntrC)	525	830
KO:K13599	two-component system, NtrC family, nitrogen regulation response regulator NtrX (ntrX)	480	443
KO:K04751	nitrogen regulatory protein P-II 1 (glnB)	282	383
KO:K04488	nitrogen fixation protein NifU and related proteins (iscU, nifU)	318	341
KO:K05521	ADP-ribosyl-[dinitrogen reductase] hydrolase [EC:3.2.2.24] (draG)	90	71
KO:K02591	nitrogenase molybdenum-iron protein beta chain [EC:1.18.6.1] (nifK)	7	16
KO:K02585	nitrogen fixation protein NifB (nifB)	6	15
KO:K02586	nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1] (nifD)	3	14
KO:K02587	nitrogenase molybdenum-cofactor synthesis protein NifE (nifE)	4	13

end reads, where > 99% of the total were identified as protein-coding genes (Table 3). Functional profiling using genes with KEGG orthologs (KO) revealed KO terms that were associated with basic cellular functions such as DNA replication, transcription and signal transduction (Table 4). Also, genes involved in nitrogen fixation were detected, demonstrating that the activity of microbial communities could be predicted with metagenomic profiling using shotgun metagenomics sequencing.

Summary

Shotgun metagenomics sequencing for soil metagenomics analysis offers many advantages over amplicon-based sequencing methods, including, but not limited to, robust taxonomic classification at the species level, *de novo* metagenome assembly, and gene detection. From soil gDNA extraction to data analysis, the workflow can be completed in less than three days. The Nextera DNA Flex Library Preparation Kit offers a simple solution for library construction, generating highly uniform libraries for even genome coverage with a wide range of DNA input. The accessibility of analytical tools via BaseSpace Sequence Hub allows for cost-effective, simple, and efficient processing of high-throughput data for a diverse array of metagenomic applications.

Learn more

Learn more about the Nextera DNA Flex Library Prep Kit at www.illumina.com/nextera-dna-flex

Learn more about IDT for Illumina - Nextera DNA UD Indexes at www.support.illumina.com/sequencing/sequencing_kits/nextera-dna-flex-kit/documentation.html

Learn more about BaseSpace Sequence Hub at www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub.html

References

1. Hillmann B, Al-Ghalith GA, Shields-Cutler RR, et al. Evaluating the information content of shallow shotgun metagenomics. *mSystems*. 2018;3(6):e00069–18.
2. Fierer N, Breitbart M, Nulton J, et al. Metagenomic and Small-Subunit rRNA Analyses Reveal the Genetic Diversity of Bacteria, Archaea, Fungi, and Viruses in Soil. *Appl Environ Microbiol*. 2007;73:7059–66.
3. Lüneberg K, Schneider D, Siebe C, Daniel R. Drylands soil bacterial community is affected by land use change and different irrigation practices in the Mezquital Valley, Mexico. *Scientific Reports*. 2018;8:1413.
4. DNeasy PowerSoil Kit, QIAGEN (2017). DNeasy PowerSoil Kit Handbook. Accessed December 18, 2018.
5. Fragment Analyzer, Agilent Technologies (2018). NGS Analysis. Accessed December 19, 2018.
6. Illumina (2017). Nextera DNA Flex Library Preparation Kit Data Sheet. Accessed December 18, 2018.
7. Kraken2 Metagenomics. www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/kraken-metagenomics.html. Accessed February 2019.
8. FASTQ Toolkit. www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/fastq-toolkit.html. Accessed December 18, 2018.
9. Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. *Nature*. 2011;473:174–180.
10. SPAdes Genome Assembler. www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/algorithmic-biology-lab-spades-genome-assembler.html. Accessed December 18, 2018
11. JGI Gold. gold.jgi.doe.gov. Accessed December 18, 2018
12. JGI IMG/M. img.jgi.doe.gov. Accessed December 18, 2018.
13. Breitweiser FP, Salzberg SL. Pavian: Interactive Analysis of metagenomics data for microbiomics and pathogen identification. *bioRxiv*. 2016;DOI:10.1101/084715.
14. Farmer RM, Tabita FR. Phosphoribulokinase mediates nitrogenase-induced carbon dioxide fixation gene repression in Rhodobacter sphaeroides. *Microbiology*. 2015;161:2184–2191.
15. Wongdee J, Boonkerd N, Teaumroong N, Tittabutr P, Giraud E. Regulation of Nitrogen Fixation in Bradyrhizhobium sp. Strain DOA9 Involves Two Distinct NifA Regulatory Proteins That Are Functionally Redundant During Symbiosis but Not During Free-Living Growth. *Front Microbiol*. 2018;9:1644.
16. Klironomos JN, McCune J, Hart M, Neville J. The influence of arbusculalr mycorrhizae on the relationship between plant diversity and productivity. *Ecol Lett*. 2000;3:137–141.
17. Ashelford KE, Day MJ, Fry JC. Elevated Abundance of Bacteriophage Infecting Bacteria in Soil. Appl Environ Microbiol. *Appl. Environ. Micobiol*. 2003;69:285–289.
18. Soil Quality Information Sheet. www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_050947.pdf. Accessed December 20, 2018.
19. Chen IA, Chu K, Palaniappan K, et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res*. 2019:47(D1):D666–D677.

Illumina, Inc. • 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.