

Terabase Metagenome Sequencing of Grassland Soil Microbiomes



Nelson WC, Anderson LN, Wu R, McDermott JE, Bell SL, Jumpponen A, Fansler SJ, Tyrrell KJ, Farris Y, Hofmockel KS, Jansson JK. 2020. Terabase metagenome sequencing of grassland soil microbiomes. *Microbiol Resour Announc* 9: e00718-20. <https://doi.org/10.1128/MRA.00718-20>.

ABSTRACT

As part of the Pacific Northwest National Laboratory's (PNNL) Science Focus Area program we are investigating the impact of environmental change on microbial community function in grassland soils. To enable a comprehensive survey of the metabolic potential of complex soil microbiomes, we performed ultra-deep metagenome sequencing, collecting >1 terabase of sequence data from grassland soils representing different precipitation regimes. Intermediate precipitation regime soil (rain-fed and irrigated agriculture) was collected from Konza Prairie Biological Station (KPBS), property of The Nature Conservancy, and managed by Kansas State University's Division of Biology.

METAGENOME BIOPROJECT ACCESSION

[KS-TmG.1.0](#)

FIELD SITE

Konza Prairie Biological Station (KPBS)
Manhattan, KS, USA

KPBS is located in the Flint Hills of northeastern Kansas on a 3,487-hectare native tallgrass prairie preserve. Landscape remains undisturbed by agriculture and considered the largest remaining area of unplowed tallgrass prairie in North America. KPBS is dominated by perennial C4 grasses and generally characterized by shallow soils overlaying chert-bearing limestones and shales. HQA and WA are burned annually, while site 4A is burned at 4-year intervals. KPBS is a Long-Term Ecological Research (LTER) site. Collection locations differed in their initial gravimetric water contents and soil texture (% water/clay) and located adjacently to the Rainfall Manipulation Plots (RaMPs).

BIOSAMPLE RECORD

Identifiers: BioSample Name: KS-TmG
Assembly Name: KS-TmG.1.0
Pooled Lab IDs: KS-1, KS-2, KS-3, KS-4, KS-5, KS-6, KS-9

Organism: soil metagenome (unclassified sequence)

Package: [MIMS: metagenome/environmental_soil; version 5.0](#)

Attributes:

collection date	2017-10-24
depth	0-20 cm
elevation	326.5 m
broad-scale environmental context	ENVO:01000177 (grassland biome)
local-scale environmental context	ENVO:00005750 (grassland soil)
environmental medium	ENVO:00001998 (soil)
geographic location	GAZ:00003937:GAZ_00004435:GAZ_00193277
latitude and longitude	39.10 N 96.61 W
annual and seasonal precipitation	835 cm
annual and seasonal temperature	-9°C - 3°C; 20°C - 33°C
pH:	6.7
total carbon (C-N):	12-18
Soil Type:	Irwin series
Soil Map:	Mollisol (ustolls)
Soil Class:	Pachic Argiustolls (fine, mixed, superactive, mesic)

Description: Complete Terabase metagenome of grassland soil microbiome collections from KPBS field site in Manhattan, KS.

BioProject: PNNL Soil Microbiome SFA 'Phenotypic Response of the Soil Microbiome to Environmental Perturbations' (70880)

METAGENOMIC STRATEGY

DNA Isolation Method: [DNeasy PowerSoil Kit](#)

Sequencing Platform: Illumina HiSeq X (pyrosequencing)

Assembly Method: MetaHipMer (version 1.0-438-g461a43b8-20190224)

Assembly Platform: <https://docs.nersc.gov/systems/cori/>

FINAL ASSEMBLY STATISTICS

Minimum Scaffold Length	Number of Scaffolds	Number of Contigs	Total Scaffold Length	Total Contig Length	Contig Coverage
All	68,100,771	68,161,098	25,868,388,303	25,866,073,288	99.99%
100	68,100,769	68,161,096	25,868,388,249	25,866,073,234	99.99%
250	68,100,766	68,161,093	25,868,388,046	25,866,073,031	99.99%
500	42,051,390	42,101,741	20,113,180,984	20,111,185,166	99.99%
1 KB	10,540,909	10,545,113	9,662,842,837	9,662,689,987	100.00%
2.5 KB	2,408,188	2,408,616	4,255,406,769	4,255,392,703	100.00%
5 KB	304,736	304,760	1,283,016,452	1,283,015,708	100.00%
10 KB	56,657	56,659	465,712,496	465,712,386	100.00%
25 KB	9,630	9,630	158,187,379	158,187,379	100.00%
50 KB	876	876	38,017,347	38,017,347	100.00%
100 KB	158	158	14,441,719	14,441,719	100.00%
250 KB	35	35	6,553,957	6,553,957	100.00%
500 KB	4	4	1,531,710	1,531,710	100.00%

DATA PACKAGE CONTENTS

MIMS Environmental/Metagenome Standard Information (metadata)

- [KS-TmG.1.0_MIMS.me.soil.5.0 \(.xlsx\)](#)

Environmental metagenome sequence (unclassified sequences) package include attributes defined by the Genome Standards Consortium (GSC) to formally describe and standardize sample metadata.

KPBS_KS-TmG.1.0_PlotMap.pdf

Field site plot location blocks and GPS coordinate information for collected soil cores.

Raw Metagenomic Data

- [RawReads_KS-TmG.1.0 \(674.53 GB; 14 items\)](#)

Name format:

<field site>-<sequencing lane>_<paired ends (forward and reverse) 1/2>.fastq.gz

KS-1_R1.fastq.gz
KS-1_R2.fastq.gz
KS-2_R1.fastq.gz
KS-2_R2.fastq.gz
KS-3_R1.fastq.gz
KS-3_R2.fastq.gz
KS-4_R1.fastq.gz
KS-4_R2.fastq.gz
KS-5_R1.fastq.gz
KS-5_R2.fastq.gz

KS-6_R1.fastq.gz
KS-6_R2.fastq.gz
KS-9_R1.fastq.gz
KS-9_R2.fastq.gz

Metagenomic Assembly Files

- [MetaHipMer_Assembly_KS-TmG.1.0 \(27.45 GB; 9 items\)](#)

File format:

.cov (average coverage across scaffolds)

KS-1.cov

KS-2.cov

KS-3.cov

KS-4.cov

KS-5.cov

KS-6.cov

KS-9.cov

KS-TmG_scaffold.fasta (assembled contigs)

KS.assembly.README (assembly parameters)

CDS and Functional Annotation Files

- [FunctionalAnnotation_KS-TmG.1.0 \(817.3 MB; 2 items\)](#)

KS-TmG.scaffold_2500.gff (annotation of biological sequences)

KS-TmG.scaffold_2500.prodigal.faa (predicted protein sequences)

DATA TAGS

Terabase metagenome, metapenome, soil microbiome, SFA, Kansas, KS, KPBS, soils, ecosystem, irrigated agriculture, marginal land, moisture regime, grasslands, environmental perturbation, LTER, intermediate moisture, rain-fed, microbial community, earth systems science, biology, data package, metadata, Illumina HiSeq X, MetaHipMer, BBTools, Prodigal, NERSC, cori platform, raw reads, assembly, protein sequence predictions, EggNOG, MIMS.soil.5.0

Acknowledgements

This research was supported by the Department of Energy Office of Biological and Environmental Research (BER) and is a contribution of the Scientific Focus Area “Phenotypic response of the soil microbiome to environmental perturbations” (70880). and the EMSL/JGI FICUS award ‘Phenotypic Response of the Soil Microbiome to Environmental Perturbations’ (51227). PNNL is operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RLO1830. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

Address Correspondence

[Janet K. Jansson](#)

Biological Systems Science
Lab Fellow, Biologist
Pacific Northwest National Laboratory
PO Box 999
MSIN: J4-18
Richland, WA 99352
janet.jansson@pnnl.gov