

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. Arid moisture regime soil (reliant on irrigated agriculture) was collected from Washington State University's Irrigated Agriculture Research and Extension Center (IAREC).

METAGENOME BIOPROJECT ACCESSION

[WA-TmG.1.0](#)

FIELD SITE

Washington State University Irrigated Agriculture Research and Extension Center (IAREC)
Prosser, WA, USA

Research and Extension Center (IAREC) field station operated by Washington State University in Washington state. Soil collections consist of two subsample variety plots (H40; 60 plots per block) organized in replicate of three locations within a ~0.5-acre (planted in May of 2018) and taken near a Hopps field east of an irrigation canal with lands sloping gently towards the south. Landscape consists of excessively drained, coarse-textured soils formed in windblown sands derived from granite, basalt, and quartzite on dunelike terraces. Plant cover consists of annual weeds, sagebrush, and wheatgrass adapted to dry northern latitudes.

BIOSAMPLE RECORD

Identifiers:	BioSample Name: WA-TmG Assembly Name: WA-TmG.1.0 Pooled Lab IDs: WA-2, WA-3, WA-4, WA-5	
Organism:	soil metagenome (unclassified sequence)	
Package:	MIMS: metagenome/environmental, soil; version 5.0	
Attributes:		
	collection date	2017-10-24T19:30:00-21:30:00 UTC
	depth	0-20 cm
	elevation	300 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:22223879
	latitude and longitude	46.25 N 119.73 W
	annual and seasonal precipitation	180 cm
	annual and seasonal temperature	8°C - 12°C
	pH:	8.0
	total organic carbon:	3.7%
	total nitrogen:	22.5 ppm
	Soil Type:	Warden series
	Soil Map:	Aridisols (cambids)
	Soil Class:	Xeric Haplocambids (coarse-silty, mixed, superactive, mesic)
Description:	Complete Terabase metagenome of grassland soil microbiome collections from IAREC field site in Prosser, WA.	
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 1.0-378-gce5e92d1)
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	83,651,096	83,671,102	30,492,625,212	30,491,910,908	100.00%
100	83,651,095	83,671,101	30,492,625,195	30,491,910,891	100.00%
250	83,651,095	83,671,101	30,492,625,195	30,491,910,891	100.00%
500	50,860,947	50,877,895	23,248,656,881	23,248,036,071	100.00%
1 KB	12,026,799	12,027,771	10,369,542,312	10,369,509,621	100.00%
2.5 KB	2,410,516	2,410,584	4,008,558,856	4,008,556,380	100.00%
5 KB	241,471	241,472	989,234,018	989,233,971	100.00%
10 KB	38,896	38,896	329,008,143	329,008,137	100.00%
25 KB	6,624	6,624	118,954,291	118,954,285	100.00%
50 KB	777	777	37,611,639	37,611,633	100.00%
100 KB	189	189	17,947,040	17,947,034	100.00%
250 KB	46	46	8,420,890	8,420,890	100.00%

DATA PACKAGE CONTENTS

MIMS Environmental/Metagenome Standard Information (metadata)

- [WA-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.

IAREC_WA-TmG.1.0_PlotMap.pdf

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

Raw Metagenomic Data

- [RawReads_WA-TmG.1.0 \(689.67 GB; 8 items\)](#)

Name format:

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

WA-2_R1.fastq.gz

WA-2_R2.fastq.gz

WA-3_R1.fastq.gz

WA-3_R2.fastq.gz

WA-4_R1.fastq.gz

WA-4_R2.fastq.gz

WA-5_R1.fastq.gz

WA-5_R2.fastq.gz

Metagenomic Assembly Files

- [MetaHipMer_Assembly_WA-TmG.1.0 \(32.32 GB; 6 items\)](#)

File format:

.cov (average coverage across scaffolds)

WA-2.cov

WA-3.cov

WA-4.cov

WA-5.cov

WA-TmG_scaffold.fasta (assembled contigs)

WA.assembly.README (assembly parameters)

CDS and Functional Annotation Files

- [FunctionalAnnotation_WA-TmG.1.0 \(1.7 GB; 3 items\)](#)

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

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

WA-TmG.scaffold_2500.prodigal.ffn (gene coding regions)

DATA TAGS

Terabase metagenome, metaphenome, soil microbiome, SFA, Washington, WSU Prosser IAREC, WA, arid soils, ecosystem, irrigated agriculture, marginal land, moisture regime, grasslands, environmental perturbation, microbial community, earth systems science, biology, data package, metadata, Illumina HiSeq X, MetaHipMer, BBTtools, 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