Microbiology Resource Announcements

| Environmental Microbiology | Announcement

Whole community shotgun metagenomes of two biological soil crust types from the Mojave Desert

Thuy M. Nguyen,[1] Nuttapon Pombubpa,[2] Marcel Huntemann,[3] Alicia Clum,[3] Brian Foster,[3] Bryce Foster,[3] Simon Roux,[3] Krishnaveni Palaniappan,[3] Neha Varghese,[3] Supratim Mukherjee,[3] T. B. K. Reddy,[3] Chris Daum,[3] Alex Copeland,[3] I-Min A. Chen,[3] Natalia N. Ivanova,[3] Nikos C. Kyrpides,[3] Miranda Harmon-Smith,[3] Emiley A. Eloe-Fadrosh,[3] Nicole Pietrasiak,[4] Jason E. Stajich,[2] Erik F. Y. Hom[1]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We present six whole community shotgun metagenomic sequencing data sets of two types of biological soil crusts sampled at the ecotone of the Mojave Desert and Colorado Desert in California. These data will help us understand the diversity and function of biocrust microbial communities, which are essential for desert ecosystems.

KEYWORDS metagenomics, soil crusts

Biological soil crusts (BSCs) cover approximately 30% of the global dryland surface area (1) and carry out important ecological services (2). As ecosystem pioneers, BSCs colonize bare soil surfaces in low-water, extreme-temperature environments. However, the underlying mechanisms for how microbes give rise to the structure and function of biocrusts and how they interact to provide ecosystem services are still not well understood.

Dry light cyanobacterial/algal crust (LAC) and cyano lichen crust (CLC) samples (~5 cm^2 in size) were collected at Sheephole Valley Wilderness within the ecotone of the Mojave Desert and Colorado Desert of San Bernardino County, CA, on 9 September 2018 with each of three replicate samples taken 250 m apart from each other at GPS location 34.1736 N, 115.3888 W, and placed in sealed glass mason jars (cushioned by paper towels) for transport back to the laboratory. One hundred milligrams of each sample were randomly subsampled and used for DNA extraction using a MoBio PowerSoil DNA Isolation Kit (12888-50; QIAGEN, Germantown, MD) following the manufacturer's instructions and sent to the DOE Joint Genome Institute (JGI) for library construction and sequencing. One hundred nanograms of extracted DNA were sheared to a target length of 656 bp using a Covaris LE220 Focused-ultrasonicator (Covaris LLC, Woburn, MA) and size selected using SPRI magnetic beads (Beckman Coulter Life Sciences, Indianapolis, IN). DNA fragments were treated with end-repair, A-tailing, and ligated to Illumina compatible adapters (Integrated DNA Technologies, Inc., Coralville, IA) using a KAPA-Illumina library kit (Roche Diagnostics Corporation, Indianapolis, IN). Libraries were quantified using a KAPA Biosystems next-generation sequencing library qPCR kit and a Roche LightCycler 480 real-time PCR instrument (Roche Diagnostics Corporation, Indianapolis, IN).

Sequencing was performed on an Illumina NovaSeq sequencer (Illumina, Inc., San Diego, CA) using XP V1 reagent kits, S4 flowcells, and a 2 × 151 indexed recipe. Reads were processed with the standard JGI Metagenome workflow (3). Briefly, BBDuk version 38.79 from the BBTools package (https://jgi.doe.gov/data-and-tools/bbtools/) was used to process raw reads, which were then corrected using BBCMS version 38.34 (k-mer count ≥2, high-count fraction ≥0.6) (4) and assembled using metaSPAdes version 3.13.0 (metagenome flag, no error correction, k-mer sizes of 33, 55, 77, 99, and 127) (5) and

Editor Frank J. Stewart, Montana State University, Bozeman, Montana, USA

Address correspondence to Erik F. Y. Hom, erik@fyhom.com.

Thuy M. Nguyen and Nuttapon Pombubpa contributed equally to this article. NP performed the sampling and DNA isolation. TMN performed the analysis and wrote the manuscript.

The authors declare no conflict of interest.

See the funding table on p. 4.

Received 13 October 2023
Accepted 23 January 2024
Published 8 February 2024

TABLE 1 Accession numbers and characteristics of metagenome data from two biocrust types from the Sheephole Valley Wilderness (Mojave Desert, CA)[a]

Metagenome	NCBI BioSample ID	NCBI BioProject ID	No. of filtered reads	Assembly BioSample ID	No. of contigs	No. of assembled (150 bp) reads	Metagenome assembly length (bp)	Metagenome coverage	N50 (bp)	Max contig length (kb)
LAC replicate 1	SAMN14510913	PRJNA620793	215,587,996	PRJNA1021617	2,550,257	166,293,935	1,742,734,840	14.3×	458,966	651.104
				PRJNA1021621	*2,332,566*	*170,423,036*	*1,868,428,908*	*13.7×*	*441,759*	*546.176*
LAC replicate 2	SAMN14510861	PRJNA620794	219,555,444	PRJNA1021629	2,663,237	162,376,141	1,700,000,691	14.3×	513,765	1000.405
				PRJNA1021637	*2,365,106*	*166,589,100*	*1,813,819,568*	*13.8×*	*479,343*	*1000.030*
LAC replicate 3	SAMN14510934	PRJNA620795	208,714,792	PRJNA1021639	2,353,418	158,432,369	1,571,681,495	15.1×	415,201	1301.000
				PRJNA1021640	*2,177,883*	*162,939,834*	*1,713,465,883*	*14.3×*	*416,349*	*541.813*
CLC replicate 1	SAMN14510746	PRJNA620790	191,564,444	PRJNA1021643	2,437,969	134,704,800	1,670,634,364	12.1×	422,767	609.588
				PRJNA1021649	*2,319,548*	*139,237,529*	*1,835,861,038*	*11.4×*	*453,111*	*169.489*
CLC replicate 2	SAMN14510985	PRJNA620791	200,791,158	PRJNA1021657	2,812,278	142,285,895	1,913,908,387	11.2×	508,862	756.686
				PRJNA1021659	*2,605,836*	*146,893,754*	*2,049,607,909*	*10.8×*	*526,639*	*384.386*
CLC replicate 3	SAMN14510745	PRJNA620792	209,432,468	PRJNA1021660	3,121,332	142,063,838	1,959,209,449	10.9×	649,941	949.882
				PRJNA1021664	*2,934,558*	*147,954,872*	*2,157,330,921*	*10.3×*	*653,857*	*345.113*

[a]All the contigs are ≥0.1 kb. Assembly statistics for metaSPAdes and MEGAHIT assemblies are listed on the top and bottom (in italics) of each row, respectively.

MEGAHIT version 1.2.9 with default settings (6). Processed reads were mapped to the final assembly, and coverage information was generated using BBMap version 38.34 (4).

Approximately 98% of raw reads passed trimming/quality control filters, and over 70% of processed reads were assembled into contigs. On average, more reads were assembled into contigs using MEGAHIT (75.8%) than metaSPAdes (72.7%) with an estimated metagenome coverage of ~13×. MetaSPAdes assemblies were more compact in terms of total assembly length, with generally longer maximum contig lengths than those of MEGAHIT, although MEGAHIT generated fewer contigs than metaSPAdes. The JGI Integrated Microbial Genomes analysis pipeline (https://genome.jgi.doe.gov/portal/ProMicSoilCrusts/ProMicSoilCrusts.info.html) revealed few eukaryotic microbes, likely due to poor DNA extraction yields from these microbes with the conventional commercial soil DNA isolation kit used. Nevertheless, these data can provide useful initial insights into the prokaryotic diversity of LAC and CLC biocrust types.

ACKNOWLEDGMENTS

We thank the BLM Needles CA office for their assistance with permitting at the Sheephole Valley Wilderness and Dr. Hung Phan (Iowa State) for help with cleaning and submitting data to NCBI.

This work was performed and supported in part by the following: the Facilities Integrating Collaborations for User Science (FICUS) program (proposal: https://doi.org/10.46936/fics.proj.2018.50356/60000035) and used resources at the DOE Joint Genome Institute (JGI) (https://ror.org/04xm1d337) and the National Energy Research Scientific Computing Center (NERSC) (https://ror.org/05v3mvq14), which are DOE Office of Science User Facilities operated under contract no. DE-AC02-05CH11231; Bureau of Land Management (BLM) Cooperative Agreement L15AC00153 to N. Pietrasiak and BLM permit number 6850-CAD0000.06 to N. Pietrasiak and J.E.S.; the U.S. Department of Agriculture, National Institute of Food and Agriculture Hatch project CA-R-PPA-211–5062-H to N. Pombubpa and J.E.S.; a Royal Thai Government Scholarship to N. Pombubpa; and NSF GoLife grant DEB-1541538 and CAREER grant DEB-1846376 to E.F.Y.H. J.E.S. is a CIFAR fellow in the Fungal Kingdom: Threats and Opportunities program. This is UM's Center for Biodiversity and Conservation Research Publication No. 40.

AUTHOR AFFILIATIONS

[1]Department of Biology and Center for Biodiversity and Conservation Research, University of Mississippi, University, Mississippi, USA
[2]Department of Microbiology and Plant Pathology, University of California, Riverside, California, USA
[3]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[4]School of Life Sciences, University of Nevada-Las Vegas, Las Vegas, Nevada, USA

PRESENT ADDRESS

Nuttapon Pombubpa, Department of Microbiology, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

AUTHOR ORCIDs

Thuy M. Nguyen http://orcid.org/0000-0002-0164-5340
Nuttapon Pombubpa http://orcid.org/0000-0003-3385-5331
Marcel Huntemann http://orcid.org/0000-0002-1284-3748
Alicia Clum http://orcid.org/0000-0002-5004-3362
Simon Roux http://orcid.org/0000-0002-5831-5895
Neha Varghese http://orcid.org/0000-0003-0580-0675
Supratim Mukherjee http://orcid.org/0000-0002-6322-2271

T. B. K. Reddy http://orcid.org/0000-0002-0871-5567
Chris Daum http://orcid.org/0000-0003-3895-5892
Alex Copeland http://orcid.org/0000-0002-3971-5439
I-Min A. Chen http://orcid.org/0000-0003-2026-9798
Natalia N. Ivanova http://orcid.org/0000-0002-5802-9485
Nikos C. Kyrpides http://orcid.org/0000-0002-6131-0462
Miranda Harmon-Smith http://orcid.org/0000-0002-3980-1820
Emiley A. Eloe-Fadrosh http://orcid.org/0000-0002-8162-1276
Nicole Pietrasiak http://orcid.org/0000-0003-4636-8006
Jason E. Stajich http://orcid.org/0000-0002-7591-0020
Erik F. Y. Hom http://orcid.org/0000-0003-0964-0031

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	DEB-1541538, DEB-1846376	Erik F. Y. Hom
U.S. Department of Energy (DOE)	FICUS-50356	Nicole Pietrasiak Jason E. Stajich Erik F. Y. Hom
Joint Genome Institute (JGI)	DE-AC02-05CH11231	Marcel Huntemann Alicia Clum Brian Foster Bryce Foster Simon Roux Krishnaveni Palaniappan Neha Varghese Supratim Mukherjee T. B. K. Reddy Chris Daum Alex Copeland I-Min A. Chen Natalia N. Ivanova Nikos C. Kyrpides Miranda Harmon-Smith Emiley A. Eloe-Fadrosh
National Energy Research Scientific Computing Center (NERSC)	DE-AC02-05CH11231	Marcel Huntemann Alicia Clum Brian Foster Bryce Foster Simon Roux Krishnaveni Palaniappan Neha Varghese Supratim Mukherjee T. B. K. Reddy Chris Daum Alex Copeland I-Min A. Chen Natalia N. Ivanova

Funder	Grant(s)	Author(s)
		Nikos C. Kyrpides
		Miranda Harmon-Smith
		Emiley A. Eloe-Fadrosh
U.S. Department of Agriculture (USDA)	CA-R-PPA-211-5062-H	Nuttapon Pombubpa
		Jason E. Stajich
Royal Thai Government	Scholarship	Nuttapon Pombubpa

AUTHOR CONTRIBUTIONS

Thuy M. Nguyen, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing | Nuttapon Pombubpa, Investigation, Methodology | Marcel Huntemann, Formal analysis, Methodology | Alicia Clum, Methodology, Supervision | Brian Foster, Formal analysis, Methodology | Bryce Foster, Methodology | Simon Roux, Supervision, Writing – review and editing, Methodology | Krishnaveni Palaniappan, Methodology | Neha Varghese, Formal analysis, Methodology | Supratim Mukherjee, Data curation | T. B. K. Reddy, Data curation | Chris Daum, Methodology, Supervision | Alex Copeland, Methodology, Supervision | I-Min A. Chen, Methodology, Supervision | Natalia N. Ivanova, Methodology, Supervision | Nikos C. Kyrpides, Methodology, Supervision | Miranda Harmon-Smith, Project administration | Emiley A. Eloe-Fadrosh, Supervision, Methodology | Nicole Pietrasiak, Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – review and editing | Jason E. Stajich, Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – review and editing | Erik F. Y. Hom, Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Raw sequencing data and metagenome assemblies are available at the NCBI using the hyperlinked BioSample and BioProject ID numbers that are listed in Table 1. Data are also available from JGI's genome portal (https://genome.jgi.doe.gov/portal/ProMicSoilCrusts/ProMicSoilCrusts.info.html) and GOLD database (https://gold.jgi.doe.gov/study?id=Gs0142145).

REFERENCES

1. Guang S, Ying Z, Haotian Y, Xinrong L. 2023. Biocrust mediates the complexity and stability of bacterial networks in both biocrust and subsoil layers in the Tengger Desert. Plant Soil 490:217–237. https://doi.org/10.1007/s11104-023-06071-x
2. Tian C, Pang J, Bu C, Wu S, Bai H, Li Y, Guo Q, Siddique KHM. 2023. The microbiomes in lichen and moss biocrust contribute differently to carbon and nitrogen cycles in arid ecosystems. Microb Ecol 86:497–508. https://doi.org/10.1007/s00248-022-02077-7
3. Clum A, Huntemann M, Bushnell B, Foster B, Foster B, Roux S, Hajek PP, Varghese N, Mukherjee S, Reddy TBK, Daum C, Yoshinaga Y, O'Malley R, Seshadri R, Kyrpides NC, Eloe-Fadrosh EA, Chen I-MA, Copeland A, Ivanova NN. 2021. DOE JGI metagenome workflow. mSystems 6:e00804-20. https://doi.org/10.1128/mSystems.00804-20
4. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory LBNL-7065E.
5. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116
6. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033