

Metagenome Sequencing of the Prokaryotic Microbiota of the Hypersaline and Meromictic Soap Lake, Washington

Erik R. Hawley^a, Matthias Hess^{a,b,c,d}

Washington State University Tri-Cities, Richland, Washington, USA^a; DOE Joint Genome Institute, Walnut Creek, California, USA^b; Washington State University, Pullman, Washington, USA^c; Pacific Northwest National Laboratory, Chemical & Biological Process Development Group, Richland, Washington, USA^d

Soap Lake is a small saline lake in central eastern Washington that is sharply stratified into two layers. In addition to being highly alkaline (~pH 10), Soap Lake also contains high concentrations of sulfide. Here, we report the community profile of the prokaryotic microbiota associated with Soap Lake surface water.

Received 12 December 2013 Accepted 19 December 2013 Published 23 January 2014

Citation Hawley ER, Hess M. 2014. Metagenome sequencing of the prokaryotic microbiota of the hypersaline and meromictic Soap Lake, Washington. *Genome Announc*. 2(1): e01212-13. doi:10.1128/genomeA.01212-13.

Copyright © 2014 Hawley and Hess. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Matthias Hess, mhess@lbl.gov.

Soda lakes and deserts are the most stable alkaline ecosystems that occur naturally on our planet (1). The ability of sodium carbonates to maintain stable alkaline pH values between 9.5 and 10.5 makes them ideal habitats for alkaliphilic microorganisms (2). The primary production in these lakes results mostly from the metabolic activities of purple sulfur bacteria and *Cyanobacteria* (3), and the daily primary production of these ecosystems exceeds in many cases the primary production measured for eutrophic lakes (4). Cultivation-independent approaches, such as DNA sequencing, are valuable tools to obtain a better understanding of the microbial communities of these extreme environments, which is a crucial step toward elucidating the factors that underlie the exceptional productivity of these ecosystems. Here, we report the metagenome, as determined by 16S rRNA amplicon sequencing, of Soap Lake surface water. Soap Lake is a small soda lake in east central Washington with a closed basin system that is permanently stratified due to the highly saline (141 g/liter) monimolimnion and the less-saline (18 g/liter) mixolimnion (5).

A sample was collected on 1 September 2012 (global positioning system [GPS] coordinates 47.421223, -119.495255) using a sterile 50-ml sampling tube. The water temperature and pH were 20°C and 10, respectively. DNA was extracted using the FastDNA SPIN kit for soil from MP Biomedicals, according to the manufacturer's protocol. The V6 to V8 region of the bacterial 16S rRNA gene was amplified using the primer set 926f/1392r and sequenced using the Roche 454 FLX+ platform (Research and Testing Laboratory, Lubbock, TX). The raw pyrosequence reads were quality filtered and analyzed using QIIME version 1.7.0 (6). Prior to filtering, the sequencing primers and barcodes were removed, allowing 1.5 mismatches to the barcode and 2 mismatches to the primer. The sequences were removed from analysis if they contained homopolymers >6 bp, were <200 bp in length, contained a quality score <25, or if they were found to be chimeric. The sequences were clustered into operational taxonomic units (OTUs) at the 97% sequence identity level using UCLUST (7), and the most abundant sequence of each OTU was chosen as a representative. The OTU representative sequences were aligned using

PyNAST (8) and then filtered to remove common gaps. The reference sequences of each OTU were taxonomically classified using the RDP Classifier (9) with an 80% confidence rating against the Greengenes database (10).

A total of 2,702 high-quality sequences, representing 172 distinct OTUs and 13 phyla, were obtained, and a Good's coverage of 0.98 was calculated. The five most abundant phyla were *Firmicutes* (45.4%), *Proteobacteria* (15.4%), *Tenericutes* (10.5%), *Actinobacteria* (10.2%), and *Bacteroidetes* (5.1%). *Cyanobacteria* contributed 3.2% of the sequence reads. Interestingly, 80.3% of the phylogenomic data were recruited by 18 OTUs, and 41.8% were by a single OTU assigned to an uncultured microorganism classified as a member of the *Firmicutes*. This data set reveals a relatively low diversity of microorganisms capable of coping with the extreme conditions of Soap Lake.

Nucleotide sequence accession number. The DNA sequences from this metagenome project have been deposited in the NCBI Short Read Archive under the accession no. SRP033727.

ACKNOWLEDGMENT

M.H., E.R.H. and the work performed in the laboratory of M.H. were funded by Washington State University.

REFERENCES

1. Wani AA, Surakasi VP, Siddharth J, Raghavan RG, Patole MS, Ranade D, Shouche YS. 2006. Molecular analyses of microbial diversity associated with the Lonar soda lake in India: an impact crater in a basalt area. *Res. Microbiol.* 157:928–937. <http://dx.doi.org/10.1016/j.resmic.2006.08.005>.
2. Sorokin DY, Kuenen JG. 2005. Haloalkaliphilic sulfur-oxidizing bacteria in soda lakes. *FEMS Microbiol. Rev.* 29:685–702. <http://dx.doi.org/10.1016/j.femsre.2004.10.005>.
3. Jones BE, Grant WD, Duckworth AW, Owenson GG. 1998. Microbial diversity of soda lakes. *Extremophiles* 2:191–200. <http://dx.doi.org/10.1007/s007920050060>.
4. Dimitriu PA, Pinkart HC, Peyton BM, Mormile MR. 2008. Spatial and temporal patterns in the microbial diversity of a meromictic soda lake in Washington State. *Appl. Environ. Microbiol.* 74:4877–4888. <http://dx.doi.org/10.1128/AEM.00455-08>.
5. Asao M, Takaichi S, Madigan MT. 2007. *Thiocapsa imhoffii*, sp. nov., an alkaliphilic purple sulfur bacterium of the family *Chromatiaceae* from

- Soap Lake, Washington (USA). *Arch. Microbiol.* 188:665–675. <http://dx.doi.org/10.1007/s00203-007-0287-9>.
6. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7:335–336. <http://dx.doi.org/10.1038/nmeth.f.303>.
 7. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <http://dx.doi.org/10.1093/bioinformatics/btq461>.
 8. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266–267. <http://dx.doi.org/10.1093/bioinformatics/btp636>.
 9. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73:5261–5267. <http://dx.doi.org/10.1128/AEM.00062-07>.
 10. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved GreenGenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 6:610–618. <http://dx.doi.org/10.1038/ismej.2011.139>.